



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 114014

**TO: Michael Borin
Location: REM/2A55
Art Unit: 1631
Wednesday, February 11, 2004**

Case Serial Number: 09/888615

**From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203**

edward.hart@uspto.gov

Search Notes

Examiner Borin,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC-Biotech/ChemLib

114014

From: Borin, Michael
Sent: Tuesday, February 10, 2004 4:17 PM
To: STIC-Biotech/ChemLib
Subject: Search request:09/888615; protease

Examiner: M.Borin
AU: 1631

Remsen 2A55
Tel.: 20713

RECEIVED
FEB 10 2004
STIC/CHEM. LIB.
(STIC)

RE: 09/888615; protease

Please conduct search of polypeptide SEQ ID 73 against the commercial and interference protein databases.

Thank you

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: *2/11/04*
Date Completed: *2/18/04*
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: *1*
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: *058*
WWW/Internet: _____
Other (specify): _____



STIC SEARCH RESULT FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact ***the searcher or contact:***

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.



Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches. Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions .rnpm and .rnpn. Searches run against the Amino Acid Pending database produce two sets of results, with the extensions .rapm and .rapn.

Because they contain data that is confidential, the results of Pending database searches should not be left in the case.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:35:56 ; Search time 52 Seconds
(without alignments)
7959.930 Million cell updates/sec

Title: US-09-888-615-73
Perfect score: 8563
Sequence: 1 MGAKSRIGFLSYEALRRV.....DTSSMBDFSDYKVCVLQ 1604

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriophage.*
- 17: sp_archaeal.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8563	100.0	1604	Q8NFA0	Q8NFA0 homo sapien
2	6807	99.5	1274	Q8EX85	Q8EX85 homo sapien
3	4344	53.1	1406	Q8IWT4	Q8IWT4 homo sapien
4	4440	51.9	828	Q9Y591	Q9Y591 homo sapien
5	2314	34.0	1736	Q9VW49	Q9VW49 drosophila
6	2604	30.4	492	Q8N3C6	Q8N3C6 homo sapien
7	1668	19.5	354	Q95482	Q95482 homo sapien
8	1406	16.4	786	Q8I5634	Q8I5634 homo sapien
9	1233	14.4	1175	Q8WT43	Q8WT43 caenorhabdi
10	1223.5	14.4	1178	Q8WT44	Q8WT44 caenorhabdi
11	1102	12.9	952	Q9R085	Q9R085 rattus norv
12	1095.5	12.8	915	Q921M8	Q921M8 mus musculu
13	1092	12.8	973	Q8AVB6	Q8AVB6 xenopus lae
14	1086	12.7	962	Q8BTL9	Q8BTL9 mus musculu
15	1063	12.4	950	Q9PWC6	Q9PWC6 gallus gall
16	1053.5	12.3	921	Q8IUG6	Q8IUG6 homo sapien

17	1018.5	11.9	910	10	Q93Y01	Q93Y01 arabidopsis
18	1006	11.7	937	10	Q9T0B8	Q9T0B8 arabidopsis
19	1004.5	11.7	1028	10	Q9ZSBS	Q9ZSBS arabidopsis
20	981.5	11.5	887	10	Q9MAQ3	Q9MAQ3 arabidopsis
21	974.5	11.4	924	10	Q9T0B6	Q9T0B6 arabidopsis
22	967.5	11.3	928	10	Q9T0B6	Q9T0B6 arabidopsis
23	939.5	11.0	887	10	Q9ZSBS	Q9ZSBS arabidopsis
24	902.5	10.5	699	11	Q99K46	Q99K46 mus musculu
25	901.5	10.5	699	11	Q8C0C2	Q8C0C2 mus musculu
26	875.5	10.2	901	10	Q9C585	Q9C585 arabidopsis
27	836	9.8	951	10	Q8H557	Q8H557 oryza sativ
28	787.5	9.2	890	10	Q8LMT7	Q8LMT7 oryza sativ
29	756.5	8.8	900	5	Q9TYH8	Q9TYH8 caenorhabdi
30	675	7.9	1095	5	Q19132	Q19132 caenorhabdi
31	651	7.6	366	11	Q8BX45	Q8BX45 mus musculu
32	607	7.1	1272	5	Q9M117	Q9M117 drosophila
33	605.5	7.1	1332	5	Q8I3U1	Q8I3U1 plasmodium
34	601	7.0	811	11	Q922N2	Q922N2 mus musculu
35	599	7.0	1272	5	Q95SG4	Q95SG4 drosophila
36	585	6.8	1298	5	Q8I1E2	Q8I1E2 drosophila
37	569	6.6	1336	5	Q8I1A0	Q8I1A0 drosophila
38	542	6.3	98	11	Q8CCP4	Q8CCP4 mus musculu
39	488	5.7	673	11	Q8R0D3	Q8R0D3 mus musculu
40	488	5.7	1080	11	Q9WVPS	Q9WVPS mus musculu
41	485	5.7	1021	11	Q8EUM9	Q8EUM9 mus musculu
42	461	5.4	470	10	Q9FPT2	Q9FPT2 arabidopsis
43	443	5.2	465	13	Q9YH21	Q9YH21 gallus gall
44	442.5	5.2	587	13	Q9VHZ0	Q9VHZ0 gallus gall
45	433	5.1	1080	11	Q9EQU1	Q9EQU1 mus musculu

ALIGNMENTS

RESULT 1
Q8NFA0 PRELIMINARY; PRT; 1604 AA.
ID Q8NFA0
AC Q8NFA0
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Ubiquitin specific protease NYREN60.
GN NYREN60.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Paulding, A., Haber, D.,
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF332301.1; J0922.1;
DR InterPro; IPR006615; DUSP.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001125; Recoverin.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00036; ehand; 2.
DR PRINTS; PR00450; RECOVERIN.
DR SMART; SM00695; DUSP; 1.
DR SMART; SM00054; EFH; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS00973; UCH_2_3; 1.
DR PROSITE; PS00973; UCH_2_3; 1.
KW Protease.
SQ SEQUENCE 1604 AA; 181654 MW; A621F764B76321E3 CRC64;
Query Match 100.0%; Score 8563; DB 4; Length 1604;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGAKSRIGFLSYEALRRVTVDELKRLKADPKRTGSLSYMGQHCFFIREVLGDGVPKVK 60

Db 1 MGAKESRIGFISYEALRRVTDVELKELKDAFKETCGLSYVMGCHCFIREVLGDGVPPKV 60
Qy 61 AEVYCSFGTSGKLHNNLIVGLVLLTRGDKBEKAKYIFSLFSSSSGNVYIREEMRL 120
Db 61 AEVYCSFGTSGKLHNNLIVGLVLLTRGDKBEKAKYIFSLFSSSSGNVYIREEMRL 120
Qy 121 HVDGKVPDTRLKCFSGEKGNYEKFNWLFNKAFTFSRWLLSGGVYVTLTDDSDTPT 180
Db 121 HVDGKVPDTRLKCFSGEKGNYEKFNWLFNKAFTFSRWLLSGGVYVTLTDDSDTPT 180
Qy 181 FYQTLAGVTHLEESDIIIDLEKRYWLLKAQRTGRFDETEGPLVSPPIRPSISEGLFNAP 240
Db 181 FYQTLAGVTHLEESDIIIDLEKRYWLLKAQRTGRFDETEGPLVSPPIRPSISEGLFNAP 240
Qy 241 DENRDNHIDFKEISCGLSACRCRGLAERQKFCFVFDVDRDGVLSVELRDMVVALLEVW 300
Db 241 DENRDNHIDFKEISCGLSACRCRGLAERQKFCFVFDVDRDGVLSVELRDMVVALLEVW 300
Qy 301 KDNRTDDIPELHMDLSDIVEGILNAHDTTKMGHLTLEDYQIWSVKNVLANEFNLILFQVC 360
Db 301 KDNRTDDIPELHMDLSDIVEGILNAHDTTKMGHLTLEDYQIWSVKNVLANEFNLILFQVC 360
Qy 361 HIVJGLRPATPEEBGQIIRGWLRESRYGLOAGHNFIIISQWQWQWKEVYKDYANPVI 420
Db 361 HIVJGLRPATPEEBGQIIRGWLRESRYGLOAGHNFIIISQWQWQWKEVYKDYANPVI 420
Qy 421 EPSSVLNGGKYSGFATAHPMEQVEDRIGSSLSYVNTTEEFSDNISTASESETAGSGFL 480
Db 421 EPSSVLNGGKYSGFATAHPMEQVEDRIGSSLSYVNTTEEFSDNISTASESETAGSGFL 480
Qy 481 YSATPGADVCFARQHTSDNNQCLLGANGNILLHNPQPGALDQPLVQBPVATSL 540
Db 481 YSATPGADVCFARQHTSDNNQCLLGANGNILLHNPQPGALDQPLVQBPVATSL 540
Qy 541 TLEGRLKRTPOLIHGRDYEMVPPVWRYALHYWGALNLPVPIKNSKTDIPELELFR 600
Db 541 TLEGRLKRTPOLIHGRDYEMVPPVWRYALHYWGALNLPVPIKNSKTDIPELELFR 600
Qy 601 YLLFLRQPATRTOQSNINWNGVNPSPNAPLKRVLAYTCGFRMOTIKEIHYLSQRL 660
Db 601 YLLFLRQPATRTOQSNINWNGVNPSPNAPLKRVLAYTCGFRMOTIKEIHYLSQRL 660
Qy 661 IKEEDMLWLYNSNYITLDDDEHKLLEYIKIOEQLVIEVRKMSWPEEFSFTANS 720
Db 661 IKEEDMLWLYNSNYITLDDDEHKLLEYIKIOEQLVIEVRKMSWPEEFSFTANS 720
Qy 721 KIDRHKVPTEKGATGLNLTGNTCFMNSSIQCVSNTPQLTQYFISGRHLYELNRTNPIGMK 780
Db 721 KIDRHKVPTEKGATGLNLTGNTCFMNSSIQCVSNTPQLTQYFISGRHLYELNRTNPIGMK 780
Qy 781 GHMAKCYGDLVQELWSGTQKNVAPLKLRTWIAKYAPRPNFGQODSQELLAPLLDGHED 840
Db 781 GHMAKCYGDLVQELWSGTQKNVAPLKLRTWIAKYAPRPNFGQODSQELLAPLLDGHED 840
Qy 841 LNRVHEKPYVELKDSGRPDWEAAEAWNDHLNRNRSIIVVDLPHGQLRSQVCKTCGHIS 900
Db 841 LNRVHEKPYVELKDSGRPDWEAAEAWNDHLNRNRSIIVVDLPHGQLRSQVCKTCGHIS 900
Qy 901 VRFPDFNPLSLPMDSYMHEITIVIKLDGTTVPYRGLRLNMDKTYGLKKQLSDICGLN 960
Db 901 VRFPDFNPLSLPMDSYMHEITIVIKLDGTTVPYRGLRLNMDKTYGLKKQLSDICGLN 960
Qy 961 SEQLLAEVHGSNINKFPQDNQKRLSVSGFLCAFEIPVPVSPISASSPTCTDFSSSPST 1020
Db 961 SEQLLAEVHGSNINKFPQDNQKRLSVSGFLCAFEIPVPVSPISASSPTCTDFSSSPST 1020
Qy 1021 NEMFTLTNGDLPRPIIPNGMNTVVP CGTEKXFTNGMNGMPSLPDPSFTGYIIAVH 1080
Db 1021 NEMFTLTNGDLPRPIIPNGMNTVVP CGTEKXFTNGMNGMPSLPDPSFTGYIIAVH 1080
Qy 1081 RKMWRTELYFLSSOKRPSLFGMLIIVPCTVHTRKDLYDAVMTQVSRLLASPLPPEASN 1140

RESULT 2

Q9BX85
ID Q9BX85 PRELIMINARY; PRT: 1274 AA.
AC Q9BX85;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Ubiquitin specific protease.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Sha J.H.;
RT "Identification of a novel ubiquitin specific protease gene related to
RT testes development from human testes cDNA library."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
EMBL AF350251; AAK30207.1;
DR NEROPS; C19.044;
DR InterPro; IPR006615; DUSP.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00443; UCH; 1.
DR SMART; SM00695; DUSP; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
KW Protease.
SQ SEQUENCE 1274 AA; 143940 MW; CB54ECB7087DC58C CRC64;
Query Match 79.5%; Score 6807; DB 4; Length 1274;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1272; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 331 MGLTLEDYQIWSVKNVLANEFNLILFQVCHIVLGLRPATPEEBGQIIRGWLRESYGL 390
Db 1 MGLTLEDYQIWSVKNVLANEFNLILFQVCHIVLGLRPATPEEBGQIIRGWLRESYGL 60

```
QY 391 QAGHNFIIISMOWQWQKSYKVDANPVVIEPSSVLNGGKYSFGTAAPMEQVEDRIGSS 450
Db 61 QAGHNFIIISMOWQWQKSYKVDANPVVIEPSSVLNGGKYSFGTAAPMEQVEDRIGSS 120
QY 451 LSYVNTTEKFSNISTASEASSETAGSGLYSATGADYCFARQHNISDNNQCILLGANG 510
Db 121 LSYVNTTEKFSNISTASEASSETAGSGLYSATGADYCFARQHNISDNNQCILLGANG 180
QY 511 NILHLNPKPGAI DNPLOVTPQPVKATSLTEGGRLKETPOLIHDRDYEMVPEPVWAL 570
Db 181 NILHLNPKPGAI DNPLOVTPQPVKATSLTEGGRLKETPOLIHDRDYEMVPEPVWAL 240
QY 571 YHWYGANLALPRVINKSKTDIPELELFPYLLFLRQOPATRTQOSNIWVNGVPSNA 630
Db 241 YHWYGANLALPRVINKSKTDIPELELFPYLLFLRQOPATRTQOSNIWVNGVPSNA 300
QY 631 PLKGVLAYTGCFSRMQTIKEIHEYL SORLRIKEEDMRILWYNSNYLTLLDDEHKL EYL 690
Db 301 PLKGVLAYTGCFSRMQTIKEIHEYL SORLRIKEEDMRILWYNSNYLTLLDDEHKL EYL 360
QY 691 KIODEOHLVTEVRNKMSPPEMSFIANSKIDRHKVPTEKATGLSNLGNTCFVNSSIQ 750
Db 361 KIODEOHLVTEVRNKMSPPEMSFIANSKIDRHKVPTEKATGLSNLGNTCFVNSSIQ 420
QY 751 CVSNTOPLQYFTSGRHLVELNRTNPIGMKGHMAKCYGDLVQELWSGTQKVAPLKLRT 810
Db 421 CVSNTOPLQYFTSGRHLVELNRTNPIGMKGHMAKCYGDLVQELWSGTQKVAPLKLRT 480
QY 811 IAKYAPRFNGFOODSOELLAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEVAEAWND 870
Db 481 IAKYAPRFNGFOODSOELLAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEVAEAWND 540
QY 871 HLNRNSIVVDLPHGQLRSQVKCTCGHISVRDPFNF.LSLPLPMDSYMHLEITVILKDG 930
Db 541 HLNRNSIVVDLPHGQLRSQVKCTCGHISVRDPFNF.LSLPLPMDSYMHLEITVILKDG 600
QY 931 TTPVRYGLRLNMDKTYGLKQJSDICLGNSEQIILAEVHGSNIKNFPQDNQKVLRSVG 990
Db 601 TTPVRYGLRLNMDKTYGLKQJSDICLGNSEQIILAEVHGSNIKNFPQDNQKVLRSVG 660
QY 991 FLCAFEIPVPSISASSPTQTDSSSPSTNEMFTLTNGDLPRIPIFNGMNTVVPVCG 1050
Db 661 FLCAFEIPVPSISASSPTQTDSSSPSTNEMFTLTNGDLPRIPIFNGMNTVVPVCG 720
QY 1051 TEKNFTNGMNGHMPSLPDSPTGYIIAIVHRKQWRTLYFLSSQKQRP.SLFGMPLI.VPCT 1110
Db 721 TEKNFTNGMNGHMPSLPDSPTGYIIAIVHRKQWRTLYFLSSQKQRP.SLFGMPLI.VPCT 780
QY 1111 VHTRKDLYDAVMOVSRLASPLPPQASNAHQD.CDDSMGYQYPTFLRVVQKDGNSCWC 1170
Db 781 VHTRKDLYDAVMOVSRLASPLPPQASNAHQD.CDDSMGYQYPTFLRVVQKDGNSCWC 840
QY 1171 PWYRFGCKIDCEDRAFIGNAYIAVMDPTALHLYQTSQBRVVDEHESVEQSRAQA 1230
Db 841 PWYRFGCKIDCEDRAFIGNAYIAVMDPTALHLYQTSQBRVVDEHESVEQSRAQA 900
QY 1231 EPINLDSCLRAFTSEELGENEMYCSKCKTHCLATKLDLWELP.LI.IHLKRFQVNG 1290
Db 901 EPINLDSCLRAFTSEELGENEMYCSKCKTHCLATKLDLWELP.LI.IHLKRFQVNG 960
QY 1291 RWIKSQXIVFPRESFDPSPALCOHKLPTPQGDSELSPELAREVKVDAQSS 1350
Db 961 RWIKSQXIVFPRESFDPSPALCOHKLPTPQGDSELSPELAREVKVDAQSS 1020
QY 1351 AGEDVLLSKPSLSANI.ISSPKGSPSSRKSGTSCPSSKNSSPNSSPRTLGRSGRLR 1410
Db 1021 AGEDVLLSKPSLSANI.ISSPKGSPSSRKSGTSCPSSKNSSPNSSPRTLGRSGRLR 1080
QY 1411 LPQIGSKNKLSSKENIDASKENGAGQICELADALSRGHVLGGSQBELVTPQDHEVALAN 1470
Db 1081 LPQIGSKNKLSSKENIDASKENGAGQICELADALSRGHVLGGSQBELVTPQDHEVALAN 1140
```

```
QY 1471 GFLYEACNGYSGNQGSHSEBSTDDOREDTRIKPIYNLYAISCHGILGG3HYVY 1530
Db 1141 GFLYEACNGYSGNQGSHSEBSTDDOREDTRIKPIYNLYAISCHGILGG3HYVY 1200
QY 1531 AKNPCKWYCYNDSSCKELHPDIDTDSAYILFYEQGIDYAOFLPKTDGKQWADTSSMD 1590
Db 1201 AKNPCKWYCYNDSSCKELHPDIDTDSAYILFYEQGIDYAOFLPKTDGKQWADTSSMD 1260
QY 1591 EDPESDYKKYCVLQ 1604
Db 1261 EDPESDYKKYCVLQ 1274

RESULT 3
Q81WT4 PRELIMINARY; PRT; 1406 AA.
AC Q81WT4;
DP 01-MAR-2003 (TREMUREL. 23, Created)
DT 01-MAR-2003 (TREMUREL. 23, last sequence update)
DT 01-MAR-2003 (TREMUREL. 23, last annotation update)
DE Ubiquitin-specific protease USP6.
GN USP6
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Paulding C., Haber D.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY143550; AAN38838.1;
KW Protease.
SQ SEQUENCE 1406 AA; 158657 MW; D3A6822CEB441DB3 CRC64;

Query Match 53.1%; Score 4544; DB 4; Length 1406;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 853; Conservative 21; Mismatches 30; Indels 4; Gaps 1;

QY 701 EVRKNDMSWPEMSFIANSKIDRHKVPTEKATGLSNLGNTCFVNSSIQCVSNTOPLQ 760
Db 499 EVHNKMSWPEMSFIANSKIDRHKVPTEKATGLSNLGNTCFVNSSIQCVSNTOPLQ 558
QY 761 YFISGRHLYELNRTNPIGMKGHMAKCYGDLVQELWSGTQKVAPLKLRTIAKYAPRFNG 820
Db 559 YFISGRHLYELNRTNPIGMKGHMAKCYGDLVQELWSGTQKVAPLKLRTIAKYAPRFNG 618
QY 821 FQQDSQELAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEVAEAWNDHLNRNSIV 880
Db 619 FQQDSQELAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEVAEAWNDHLNRNSIV 678
QY 881 DLPHGQLRSQVKCTCGHISVRDPFNF.LSLPLPMDSYMHLEITVILKDGTPVRYGLRL 940
Db 679 DLPHGQLRSQVKCTCGHISVRDPFNF.LSLPLPMDSYMHLEITVILKDGTPVRYGLRL 738
QY 941 NMDKYTGLKQJSDICLGNSEQIILAEVHGSNIKNFPQDNQKVLRSVGFCAFEIPVP 1000
Db 739 NMDKYTGLKQJSDICLGNSEQIILAEVHGSNIKNFPQDNQKVLRSVGFCAFEIPVP 798
QY 1001 VSPISASSPTQTDSSSPSTNEMFTLTNGDLPRIPIFNGMNTVVPVCGTEKNFTNGMV 1060
Db 799 VSPISASSPTQTDSSSPSTNEMFTLTNGDLPRIPIFNGMNTVVPVCGTEKNFTNGMV 858
QY 1061 NGHMPSLSDSPFTGYIIAIVHRKQWRTLYFLSSQKQRP.SLFGMPLI.VPCTVHTRKDLYD 1120
Db 859 NGHMPSLSDSPFTGYIIAIVHRKQWRTLYFLSSQKQRP.SLFGMPLI.VPCTVHTRKDLYD 918
QY 1121 AVWQVSLASPLPPQASNAHQD.CDDSMGYQYPTFLRVVQKDGNSCWCWYRFGCK 1180
Db 919 AVWQVSLASPLPPQASNAHQD.CDDSMGYQYPTFLRVVQKDGNSCWCWYRFGCK 978
QY 1181 IDCEDRAFIGNAYIAVMDPTALHLYQTSQBRVVDEHESVEQSRAQAEPINLDSCLR 1240
Db 979 IDCEDRAFIGNAYIAVMDPTALHLYQTSQBRVVDEHESVEQSRAQAEPINLDSCLR 1038
```

```
QY 1241 APTSEELGENMYGCKKTHCLATKLDLWRLPPIIHLKRFQVNGRMVKSQKIVK 1300
Db 1039 APTSEELGESEMYGCKKTHCLATKLDLWRLPPIIHLKRFQVNDQWIKSQKIVR 1098
QY 1301 FPRESFPAFLVPRDPAALCOHKPLTPQGEISEPRILAREVKVDAQSSAGEEDVLSK 1360
Db 1099 FLRESFPAFLVPRDPAALCOHKPLTPQGEISEPRILAREVKVDAQSSAGEEDVLSK 1158
QY 1361 SPSSLSANIISPKGSPSSSRKSGTSCPSKSSPNSSPRTIGRSKRLRLPOIGSKNKL 1420
Db 1159 SPSSLSANIISPKGSPSSSRKSGTSCPSKSSPNSSPRTIGRSKRLRLPOIGSKNKP 1218
QY 1421 SSSKENLDASKENGAGCICELADALSGHVLGGSQPELVTPQDHEVALANGFLYEACG 1480
Db 1219 SSSKENLDASKENGAGCICELADALSGHVRGGSQPELVTPQDHEVALANGFLYEACG 1278
QY 1481 N----GYSNGLGNHSEESTDDOREDTIKPIYNLYAISCHSGILGGHYTYAKNPNC 1536
Db 1279 NGCGDYSNGLGNHSEESTDDOREDTIKPIYNLYAISCHSGILGGHYTYAKNPNC 1338
QY 1537 KWCYNDSSCKELHPDEIDTDSAYILFYEQOQIDYAFKPKTDGKMAWTSMDDEDFESD 1596
Db 1339 KWCYNDSSCKELHPDEIDTDSAYILFYEQOQIDYAFKPKTDGKMAWTSMDDESD 1398
QY 1597 YKXVCVLQ 1604
Db 1399 YEKYSMLQ 1406

RESULT 4
Q9V591 PRELIMINARY; PRT; 828 AA.
ID Q9V591
AC Q9V591;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NY-REN-60 antigen (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
SEQUENCE FROM N.A.
MEDLINE=99438124; PubMed=10508479;
RA Scanlan M.J., Gordan J.D., Williamson B., Stockert E., Bander N.H.,
RA Jongeneel V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,
RA Old L.J.;
RT "Antigens recognized by autologous antibody in patients with renal-
RL cell carcinoma."
RL Int. J. Cancer 83:456-464(1999).
DR EMBL; AF155116; AAD42882.1; -.
DR MEROPS; C19.044; -.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS50235; UCH_2_3; 1.
FT NON TER 1
FT NON TER 828
FT NON TER 828
SQ SEQUENCE 828 AA; 94806 MW; EEB03264D50DDCC2 CRC64;

Query Match 51.9%; Score 4440; DB 4; Length 828;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 536 KATSLTLEGRLKPTLPOLIGHRDYEMVPEPVWALVHWYCANALPRPVKSKTDIPEL 595
Db 1 KATSLTLEGRLKPTLPOLIGHRDYEMVPEPVWALVHWYCANALPRPVKSKTDIPEL 60
QY 596 ELFPYLLFLRQQPATRQTSQNIWVWNGVPSNAPLKRVLAYTGCFSRMQTIKEIHEYL 655
Db 61 ELFPYLLFLRQQPATRQTSQNIWVWNGVPSNAPLKRVLAYTGCFSRMQTIKEIHEYL 120
```

```
QY 656 SQRIRKEEDRWLWLNSENLYTLDDDEHKLBYLKIQDEQHLVIEVRNKMDSWPBMSF 715
Db 121 SQRIRKEEDRWLWLNSENLYTLDDDEHKLBYLKIQDEQHLVIEVRNKMDSWPBMSF 180
QY 716 IANSKIDRHKVPTEKATGSLNLTGNTCFNNSIQCVSNTQPLTOYFISGRHLYELNRTN 775
Db 181 IANSKIDRHKVPTEKATGSLNLTGNTCFNNSIQCVSNTQPLTOYFISGRHLYELNRTN 240
QY 776 PIGMKHMAKCYGLDVOELWSGTQKNVAPLKLAWTIKAPRNGFOODSQELLAPLLD 835
Db 241 PIGMKHMAKCYGLDVOELWSGTQKNVAPLKLAWTIKAPRNGFOODSQELLAPLLD 300
QY 836 GLHEDLNVRHEKPYVELKSDGRPDWEVAEAWNDHLRRNRSIVDLFHQQLRSQVKCKT 895
Db 301 GLHEDLNVRHEKPYVELKSDGRPDWEVAEAWNDHLRRNRSIVDLFHQQLRSQVKCKT 360
QY 896 CGHLSVRFDFPNFLSLPLPMDSYMHLEITVILKDGTTFVRYGLRLANDKVTGLKKQLSD 955
Db 361 CGHLSVRFDFPNFLSLPLPMDSYMHLEITVILKDGTTFVRYGLRLANDKVTGLKKQLSD 420
QY 956 LCGLNSQIILLAEVHGSIKNFPQDNQKRLSVSGFLCAFEIPVPVSPISASSPTQDFFS 1015
Db 421 LCGLNSQIILLAEVHGSIKNFPQDNQKRLSVSGFLCAFEIPVPVSPISASSPTQDFFS 480
QY 1016 SSPSTNEMFTLTNGDLPRDIFIPNGMPNTVPCGTEKNTGNGVGHMPSLSDSPETGY 1075
Db 481 SSPSTNEMFTLTNGDLPRDIFIPNGMPNTVPCGTEKNTGNGVGHMPSLSDSPETGY 540
QY 1076 IIAVHRKQMTELYFLSSQKRNPSLFCMPLIVPCTVHTRKKLDYDAWIOVSRLASPLPP 1135
Db 541 IIAVHRKQMTELYFLSSQKRNPSLFCMPLIVPCTVHTRKKLDYDAWIOVSRLASPLPP 600
QY 1136 QEASNHAAQDDDSMGYQYPTFLRVQKDGNSCAWCPWYRCRCCKIDCGEDRAFIGNAYI 1195
Db 601 QEASNHAAQDDDSMGYQYPTFLRVQKDGNSCAWCPWYRCRCCKIDCGEDRAFIGNAYI 660
QY 1196 AVDWDPALHLRYOTSQERVVDEHESVEQSRRQAEPINLDSCLRAFTSEELGENEMY 1255
Db 561 AVDWDPALHLRYOTSQERVVDEHESVEQSRRQAEPINLDSCLRAFTSEELGENEMY 720
QY 1256 CSKCKTHCLATKLDLWRLPPIIHLKRFQVNGRMVKSQKIVKPFRESFDSAFLVPR 1315
Db 721 CSKCKTHCLATKLDLWRLPPIIHLKRFQVNGRMVKSQKIVKPFRESFDSAFLVPR 780
QY 1316 DPALCQHKPLTPQGEISEPRILAREVKVDAQSSAGEEDVLSKSPS 1363
Db 781 DPALCQHKPLTPQGEISEPRILAREVKVDAQSSAGEEDVLSKSPS 828

RESULT 5
Q9VW49 PRELIMINARY; PRT; 1736 AA.
ID Q9VW49
AC Q9VW49;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG8334 protein.
GN CG8334.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX [1]
SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
```



```
Db 851 LLSLPLPVNYFYFVLVILLOSVPKIKYFRLNSDCKYSHLKHLSLWCSLPPNMLVC 910
QY 968 EVHGSNIKFPQDNQKRVLSVGFCAPIV-
Db 911 EWNISQIRVLNDEKLRQSAKELYVQLPEQSMRTSRNSGLSMHIEQGLKDIQRSSAL 970
QY 1000 -----PV- 1001
Db 971 ITSQDSLSLSTLTQSSSHRASSRVLNCHVSGLDVEGEAEVGTDVQCSNSNNTPIVS 1030
QY 1002 ----- 1001
Db 1031 TVSGNGSGDNQVHELLPDEAGKDELAPQRNCYSRMIYRRWFDCECEIRQESASVCF 1090
QY 1002 ---SPISASPP-----TQDSSSSTNEMFTLITNGLPRI 1036
Db 1091 SIESLSSSPNTFVHGAQAQKRVSSAKLHSTNSMSYTHSGNSMSSLTPEI 1150
QY 1037 FIP-----NGMPNTVPCGTGKNTGMVNGHM--PSL-----PDS 1070
Db 1151 PLADLEPVSRNGSGGDCSYRTSPNDSSGLSTGHTLGASLDVQAEAGNAEDHDQPDQ 1210
QY 1071 PFTG-----YIAVHRKMRTELYFLSQKRPSPFLGMLPVP 1108
Db 1211 ITTSQPETSGVYSRRSSQPPHAKGKYLVAHRKITRSDSYFLSKTRPSLFGVPLLP 1270
QY 1109 CTVHTKDKLDYDAVITQVSRLASPLP-POEASNAHQDCDDSMGYQYFPTLRVVQKDGNSC 1167
Db 1271 NSGGTHKLDYCAVILQVSELLSPLPATTEQANHAADCDLSGYDPFPTLRVAKDGLTC 1330
QY 1168 AWCPTWRFRCGCKIDCEGDRAFG-----NA----- 1193
Db 1331 AICPWSFRCGCEIRCNNDYVLOGALPPIAAASNTSTPKMAKFPSPLENRAKTPPT 1390
QY 1194 -----YIADWDPTALHRYQTSQERVVDEHESVEQSRRAQAEPTNLDSCL 1238
Db 1391 ASLSYPTTKYFEDFTALDWDPTALHRYQTSLERLWVDHETIALSREQVEPDLNHC 1450
QY 1239 LRAFTSEELGENEMYCCKTKHCLATKLDLWRLPPIIHLKRFQFVNGRWIKSKQIVK 1298
Db 1451 LRAFTSEEL--EQYHCHSKGKPKATKQIWLKPLPILIVHLKRFVNCVNGKWKQSV 1508
QY 1299 VKPRSFDSAEI--VPRDPALCOHKPLTPQDELSEPRILAREVKVYDAQSSAGEEDV 1356
Db 1509 VHPFPDDFDPTPLASVFOETIL-RHKELL-----EL-----KNDAMTWATNEV 1552
QY 1357 LLSKSPSSLSANIISPKGSPSSSRKSGTSCPS--KNSSPSSPRTLGSRKGLRLPQI 1414
Db 1553 V-----SELDEIDAPSKEVKEELPNOT-----GSTRKATASPPPT 1586
QY 1415 GSKNKLSSKENLDASKENGAGQICELADALSRGHVGGSQPELVTPQDHEVALANGFLY 1474
Db 1587 GNILROSKTK-----NAVRQRLISTSLTK--TP-----IVDGEFE 1620
QY 1475 EHEACGNGYNGOLGNHSESDTDQREDTRIKPIYNLYAISCHGILGGHVVYVAKNP 1534
Db 1621 DYH-----QRLPDDVD-----QDFRYELVAVVSHSGMLNGHHYISVASNA 1662
QY 1535 NCRWYCNDSCKEL-HPDEIDTSAVILFYEQGIDYAFPLPKTDGKKVADTSSMDEPESDYKY 1600
Db 1663 TGSWYCNDSCKEISQKVIDPSAAYLLFYERKGLDYPFLPNIBGRTLPNTASVPLEV 1722
QY 1591 EDESDYKVC 1601
Db 1723 DETEGELKKLC 1733
```

RESULT 6

Q8N3C6
ID Q8N3C6 PRELIMINARY; PRT; 492 AA.
AC Q8N3C6;
DT 01-OCT-2002 (Tremblrel. 22, Created)

01-OCT-2002 (Tremblrel. 22, Last sequence update)
01-MAR-2003 (Tremblrel. 23, Last annotation update)
GN DXFP761B049.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Amnydala;
RA Blum H., Bauersachs S., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834444; CAD39104.1; -
DR InterPro; IPR001394; UCH-2.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS50235; UCH_2_3; 1.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 492 AA; 54836 MW; C00D9438503B51E5 CRC64;
Query Match 30.4%; Score 2604; DB 4; Length 492;
Best Local Similarity 100.0%; Pred No. 2.2e-187;
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1121 AVTIQVSRLASPLPPOEASNAHQDCDDSMGYQYFPTLRVVQKDGNSCAMCPTWRFRCGCK 68
Db 9 AVTIQVSRLASPLPPOEASNAHQDCDDSMGYQYFPTLRVVQKDGNSCAMCPTWRFRCGCK 68
QY 1181 IDGEDRAFTGNAYIADVDPDPTALHRYQTSQERVVDEHESVEQSRRAQAEPTNLDSCLR 1240
Db 69 IDGEDRAFTGNAYIADVDPDPTALHRYQTSQERVVDEHESVEQSRRAQAEPTNLDSCLR 128
QY 1241 AFTSEELGENEMYCCKTKHCLATKLDLWRLPPIIHLKRFQFVNGRWIKSKQIVK 1300
Db 129 AFTSEELGENEMYCCKTKHCLATKLDLWRLPPIIHLKRFQFVNGRWIKSKQIVK 188
QY 1301 PPRSFPSAFLVPRDPALCOHKPLTPQDELSEPRILAREVKVYDAQSSAGEEDVLLSK 1360
Db 189 PPRSFPSAFLVPRDPALCOHKPLTPQDELSEPRILAREVKVYDAQSSAGEEDVLLSK 248
QY 1361 SPSSLSANIISPKGSPSSSRKSGTSCPSKNSPSSPRTLGSRKGLRLPQIGSKNKL 1420
Db 249 SPSSLSANIISPKGSPSSSRKSGTSCPSKNSPSSPRTLGSRKGLRLPQIGSKNKL 308
QY 1421 SSSKENLDASKENGAGQICELADALSRGHVGGSQPELVTPQDHEVALANGFLYSEACG 1480
Db 309 SSSKENLDASKENGAGQICELADALSRGHVGGSQPELVTPQDHEVALANGFLYSEACG 368
QY 1481 NGYNGOLGNHSESDTDQREDTRIKPIYNLYAISCHGILGGHVVYVAKNPCKWYC 1540
Db 369 NGYNGOLGNHSESDTDQREDTRIKPIYNLYAISCHGILGGHVVYVAKNPCKWYC 428
QY 1541 YNDSCKELHPDEIDTSAVILFYEQGIDYAFPLPKTDGKKVADTSSMDEPESDYKY 1600
Db 429 YNDSCKELHPDEIDTSAVILFYEQGIDYAFPLPKTDGKKVADTSSMDEPESDYKY 488
QY 1601 CVLQ 1604
Db 489 CVLQ 492
RESULT 7
Q95482
ID Q95482 PRELIMINARY; PRT; 354 AA.
AC Q95482;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE TL132 protein.
GN TL132
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RA Goulin F., Dominguez J., Draetta G.F.;
RT "CDNA cloning of an isoform of the human tre oncogene";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; A012755; CAA01071.1; --
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
SQ SEQUENCE 354 AA; 39586 MW; 7112708E118EC127 CRC64;

Query Match 19.5%; Score 1668; DB 4; Length 354;
Best Local Similarity 92.7%; Pred. No. 4 4e-117;
Matches 316; Conservative 8; Mismatches 17; Indels 0; Gaps 0;

QY 1264 LATKLDLWRLPPIIHLKRFQVNGRWIKQKIVKFPRESFDPFSAFLVPRDPALCOHK 1323

DB 14 LATKLDLWRLPPIIHLKRFQVNDQWIKQKIVKFPRESFDPFSAFLVPRDPALCOHQ 73

QY 1324 PLTPQDELSEPRILAREVKYKDAOSSAGEEDVLLSKSPSSLSANTISSPKGSPSSRS 1383

DB 74 PLTPQDELSEPRILAREVKYKDVQSLAGEEDVLLSKSPSSLSANTISSPKGSPSSRS 133

QY 1384 GTSPCKSNSSPNSSPRTILGRSGRLRFPQIGSKNKLSSKENLDASKENGAGQICELAD 1443

DB 134 GTSPCKSNSSPNSSPRTILGRSGRLRFPQIGSKNKLSSKENLDASKENGAGQICELAD 193

QY 1444 ALSRGHVGSGQBELVTPQDHEVALANGFLYEACGNGYNGOLGNHSEEDSTDQDRED 1503

DB 194 ALSRGHVGSGQBELVTPQDHEVALANGFLYEACGNGYNGOLGNHSEEDSTDQDRED 253

QY 1504 TRIKPIYNLYAISCHSGILGGHYVYAKNPCKWYCYNDSSCKELHPDEITDTSAYILF 1563

DB 254 TRIKPIYNLYAISCHSGILGGHYVYAKNPCKWYCYNDSSCKELHPDEITDTSAYILF 313

QY 1564 YEQQIDYAOFLPKDTGKMDTSSMDEDFESDYKYCVLQ 1604

DB 314 YEQQIDYAOFLPKDTGKMDTSSMDEDFESDYKYCVLQ 354

RESULT 8

Q15634 ID Q15634 PRELIMINARY; PRT; 786 AA.

AC Q15634;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Oncogene.
GN TRE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=Ewing's sarcoma;
RX MEDLINE=92228503; PubMed=1565468;
RA Nakamura T., Hillova J., Mariage-Samson R., Onno M., Huebner K.,
RA Cannizzaro L.A., Boghosian-Sell L., Croce C.M., Hill M.;
RT "A novel transcriptional unit of the tre oncogene widely expressed in
human cancer cells";
RL Oncogene 7:733-741(1992).

DR EMBL; X63546; CAA45108.1; --

DR InterPro; IPR000195; RABGAP.TBC.

DR Pfam; PF00566; TBC; 1.

DR Pfam; PF00443; UCH; 1.

DR SMART; SM00164; TBC; 1.

DR PROSITE; PS0086; TBC.RABGAP; 1.

DR PROSITE; PS00972; UCH_2_1; 1.

DR PROSITE; PS0235; UCH_2_3; 1.
SQ SEQUENCE 786 AA; 83546 MW; CC0C9280ED15DA0A CRC64;

Query Match 16.4%; Score 1406; DB 4; Length 786;
Best Local Similarity 92.7%; Pred. No. 9 4e-97;
Matches 266; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 701 EVRNKMSWPEEMSFANSKIDRHKVPTKGTGLSLNLTGTCFMSSICVSNTPQLTQ 760

DB 499 EVRNKMSWPEEMSFANSKIDRHKVPTKGTGLSLNLTGTCFMSSICVSNTPQLTQ 558

QY 761 YFISGRHLYELNRTNPTGMKGMAKCYGDIQVBLWSGTQKRVAPLKLRTWIAYAPFNG 820

DB 559 YFISGRHLYELNRTNPTGMKGMAKCYGDIQVBLWSGTQKRVAPLKLRTWIAYAPFNG 618

QY 821 FQODSOELLAFLDGLGHEDLNRVHEKPYVELKDSGRPDWEVAAEAWDNHNRNSIV 880

DB 619 FQODSOELLAFLDGLGHEDLNRVHEKPYVELKDSGRPDWEVAAEAWDNHNRNSIV 678

QY 881 DLPHGQLRSQVKCTCGHISVRFPDFFNLSPLPMDSYHLEITVVKLDTTPVRYGLRL 940

DB 679 DLPHGQLRSQVKCTCGHISVRFPDFFNLSPLPMDSYHLEITVVKLDTTPVRYGLRL 738

QY 941 NMDEKTYGLKQLSDICGLNSEQILLAEVHSGNIKPPQDNQKVRLS 987

DB 739 NMDEKTYGLKQLSDICGLNSEQILLAEVHSGNIKPPQDNQKVRLS 785

RESULT 9

Q8WT43 ID Q8WT43 PRELIMINARY; PRT; 1175 AA.

AC Q8WT43;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 133.8 kDa protein (Ubiquitin c-terminal hydrolase).
GN ZK328.1 OR CYK-3.

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";

RL Science 282:2012-2018(1998).

RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Favello A.;

RT "The sequence of C. elegans cosmid ZK328.";

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.;

RT "Direct Submission.";

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

RN [4]
RP SEQUENCE FROM N.A.

RA Kaitna S., Schnabel H., Hyman A.A., Glotzer M.;

RT "An ubiquitin C-terminal hydrolase is required to maintain osmotic
balance and execute actin-dependent processes in the early
Caenorhabditis elegans embryo.";

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; U50193; AAL32262.1; --

DR EMBL; AF469173; AAL79016.1; --

DR WormPep; ZK328.1a; CE30061.

DR InterPro; IPR006615; DUSP.

DR InterPro; IPR002048; EF-hand.

DR	PROSITE; PS00973; UCH_2_2; 1.	
DR	PROSITE; PS50235; UCH_2_3; 1.	
KW	Nuclear protein.	
SQ	SEQUENCE 915 AA; 103752 MW; 1FF0537D187D4904 CRC64;	
	Query Match	12.8%; Score 1095.5; DB 11; Length 915;
	Best Local Similarity	24.8%; Pred. No. 3.1e-73;
	Matches 318; Conservative 166; Mismatches 323; Indels 477; Gaps 37;	
QY	390 LQAGHNWFIISMQWQKWEKYVKYDANVPVIEPSSVLNKGKYGFGTAAPMEQVEDRIGS	449
Db	28 LQGAQWFLIDSRWFKQWKYVGFD	60
QY	450 SLSYVNTTEKFSNIDNSTASEASETAGSGFLYSATPGADYCFARQHNNTSDNNQCLLGAN	509
Db	61	65
QY	510 GNILLHLNPKPGALDNPQVLVTPQEVKATSLTLEGGLKRTPTLIHGRDVMYMPVPWRA	569
Db	66	103
QY	570 LYHWYGANLALPRPVIKNSKTDIPELELPR	623
Db	104 LLNWIYC-VGGQPIVRK	142
QY	624 NVPSNPAPLKRVLAYTGC-PSRMQTIKEIHYSQRLRI-KEEDMRLWLNSENLYTLTD	681
Db	143	196
QY	682 DEDHKLEVLKIQDEHVLVIEVRKNDMSWPEMSFIAN	734
Db	197 KLNTIIOAGLYQQGVLIIPQNEDEGWPPQ-SLQSNMGSGFSASYNCQEPSPSHIQGLC	255
QY	735 GLSLNGNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRTNPIGMKHMAKCYGDLVQL	794
Db	256 GLGLNGNTCFMNSALQCLSNAPTLEYFLKDEYEAENRDNPLGMKGEIAEAYAEILKQM	315
QY	795 WSGTQKNVAPLKLRTWTAKYAPRNGSQOQDSQELLAFILDLGHEDLNVRHVEPYVELKD	854
Db	316 WSGEDTHVAPRMFTQVGRFAPQSGYQOQDSQELLAFILDLGHEDLNVRHVEPYVELKD	375
QY	855 SDGRPDVEAAEAWNDHLNRNRSIVDFLPHQSLRSQVCKCTCGHISVRFDPFPNPLSLPLP	914
Db	376 ANGRPDVAVKAENHRLNDSVIVDTFHLGPKSLTLCVPECAKVSVTDFDFCYLTPLP	435
QY	915 MDSVMHLEITVILKD-GTTPVRYGLRLNMBEKYTLGLKQLSDLCLGNSEQILLAEVHGSN	973
Db	436 LKQDRIMEVFLVPADPCQRIQYRTVPLMGALSDLCALSKLSGIAAENNVVTDVYNHR	495
QY	974 I-KNFPQDNQKVLRSVSGFLCAFEIPVPSPIASSTQTDFFSSPSTNEMFTLTITNGDL	1032
Db	496 FHKIFQWDEGLSHITPRDDIPVIEV	533
QY	1033 PRPIPINGMNTVVPCTEKNFTNGVMNGHMPSLDPSPTGYTIAVHRKMRTELTYFLS	1092
Db	534	552
QY	1093 SQKNRPSLFGMLPIVPCTVH-TRKDLVDAYWIOVSR	1135
Db	553	605
QY	1136	1161
Db	606	664
QY	1162 KDGNSCAWCPVYRCRGCKIDCGEDRAPFIGNAYIAVMDPTALHLRYOTSQERVVDHSES	1231
Db	665	716
QY	1222 VEQRRQAQEPINLDSCLRAFTSBEELGENEMYYCSKTKTCLATKKLDLWRLPILIIH	1281
Db	717	776

715	Db	715	RSFLALDWDPELKGRYFDDSAADPEKHESVEY--KPPKPPFVGLKDCIELEFTEKCLGA	772
1251	QY	1251	NEMWYCSCKTCELATKKLDLWRLPPLIIHLKRFQVNGRWIKSQKIVKFPRESFDPSA	1310
773	Db	773	EDFWYCPNCKEHOQATKKLDLWLSLPVLVHLKRFYSRYMDKLDLTUVGFINDLMS-	831
1311	QY	1311	FLVPRDPALCOHKPLTPQGDLSSEPRILAREVKVKVDAQSSAGEEDVLLSKPSSLSANII	1370
832	Db	832	--IP-----	833
1371	QY	1371	SSPKGSPSSSRKSGTSCPSSKNSPNSSPRTLGRSGRLRLPQIGSKNKLSSXENLDAS	1430
834	Db	834	-----YNPVAGP-----	840
1431	QY	1431	KENGAGQICELADALSRGHVLCGSQBELVTPQDHEVALANGFLYEHCAECNGYSNGQLN	1490
841	Db	841	-----CR-----	842
1491	QY	1491	HSEEDSTDQREDTRIKPIYNLVAISCHSGILGGHYVTYAKN-PNCKWYVNDSSCKEL	1549
843	Db	843	-----YNLIANSNHVCGGCGHYTAFVKNKDDGKWIYFDDSSVSTA	883
1550	QY	1550	HPDEIDTDSAYILFYBQQ---GIDYAQFLPKTDGKKMA-----	1593
884	Db	884	CEDQIVSKAAVLYFYORQDTISGTGF--FLDRETKQGASAAATGIPLESDSDSNENDNDI	941
1594	QY	1594	ESD 1596	
942	Db	942	ENE 944	

Search completed: February 11, 2004, 15:40:20
Job time : 64 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:32:01 / Search time 53 Seconds
(without alignments)
4803.725 Million cell updates/sec

Title: US-09-888-615-73

Perfect score: 8563

Sequence: 1 MGAKESRIGFLSYEALRRV.....DTSMDEDFSDYKVCVLQ 1604

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A_Genesec19Jun03.*
2: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1980.DAT.*
3: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1981.DAT.*
4: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1982.DAT.*
5: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1983.DAT.*
6: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1984.DAT.*
7: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1985.DAT.*
8: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1986.DAT.*
9: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1987.DAT.*
10: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1988.DAT.*
11: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1989.DAT.*
12: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1990.DAT.*
13: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1991.DAT.*
14: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1992.DAT.*
15: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1993.DAT.*
16: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1994.DAT.*
17: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1995.DAT.*
18: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1996.DAT.*
19: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1997.DAT.*
20: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1998.DAT.*
21: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1999.DAT.*
22: /SIDS1/gcgdata/genesec/genesecp-emb1/AA2000.DAT.*
23: /SIDS1/gcgdata/genesec/genesecp-emb1/AA2001.DAT.*
24: /SIDS1/gcgdata/genesec/genesecp-emb1/AA2002.DAT.*
25: /SIDS1/gcgdata/genesec/genesecp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8563	100.0	1604	23	AAU82715 Amino acid sequenc
2	6826	79.7	1317	23	ABG32160 Novel human ubiqui
3	6807	79.5	1274	23	AAU75578 Human ubiquitin sp
4	5570.5	65.1	1059	24	ABU12111 Human protein mod
5	4822	56.3	1127	22	AAU40477 Human polypeptide
6	4822	56.3	1127	22	AAU40478 Human polypeptide
7	4543	53.1	1085	22	AAU38692 Human polypeptide
8	4531	52.9	1089	19	AAU82396 Human UBP protein
9	4531	52.9	1089	22	ABG23332 Novel human diagn

10	4531	52.9	1089	22	AAU38691 Human polypeptide
11	4433.5	51.8	1125	22	ABG21348 Novel human diagn
12	4328.5	50.5	1388	22	ABG21344 Novel human diagn
13	3138	36.6	582	23	AAE19177 Human protease PR
14	3019.5	35.3	1779	22	ABE71649 Drosophila melanog
15	2711	31.7	539	22	AAU87244 Novel central nerv
16	2509	29.3	1066	22	ABG24025 Novel human diagn
17	2392.5	27.9	462	21	AAU92344 Human cancer assoc
18	1801	21.0	1131	22	ABG13147 Novel human diagn
19	1801	21.0	1142	22	ABG24027 Novel human diagn
20	1750	20.4	401	22	AAU43544 Human polypeptide
21	1406	16.4	805	22	ABG24026 Novel human diagn
22	1140.5	13.3	439	22	ABG21346 Novel human diagn
23	1140.5	13.3	439	22	ABG24024 Novel human diagn
24	1081	12.6	952	21	AAU70014 Human Protease and
25	1053.5	12.3	920	23	ABP68903 Human polypeptide
26	1006	11.7	929	21	ABG31411 Arabidopsis thalia
27	1006	11.7	937	21	ABG31410 Arabidopsis thalia
28	1006	11.7	937	24	ABF59344 AAG10590 protein
29	938.5	11.0	1024	22	AAU31808 Novel human secret
30	932.5	10.9	835	21	ABG31412 Arabidopsis thalia
31	920.5	10.7	909	21	AAU31701 Arabidopsis thalia
32	920.5	10.7	914	21	AAU31700 Arabidopsis thalia
33	890	10.4	173	22	AAU87546 Novel central nerv
34	890	10.4	173	22	AAU43622 Human polypeptide
35	890	10.4	173	22	AAU19956 Novel human calcu
36	876	10.2	196	24	ABU00140 Human novel polype
37	837.5	9.8	320	22	ABG21345 Novel human diagn
38	787.5	9.2	748	21	AAU31702 Arabidopsis thalia
39	738.5	8.6	1254	22	AAU70760 S cerevisiae apopt
40	713	8.3	1372	22	AAU70881 C albicans apotos
41	665	7.8	1318	22	AAU28169 Novel human secret
42	665	7.8	1373	21	AAU42190 Human ORFX ORF1954
43	647.5	7.6	712	22	ABG00341 Novel human diagn
44	607	7.1	1096	22	ABB60222 Drosophila melanog
45	524	6.1	1123	23	ABU98133 Human PMM incyte

ALIGNMENTS

RESULT 1

AAU82715
ID AAU82715 standard; Protein; 1604 AA.
XX AC AAU82715;
XX AC

23-APR-2002 (first entry)

Amino acid sequence of novel human protease #14.

Human; protease; cancer; immune-related disorder; cardiovascular disease;
neutonal-associated disease; metabolic disorder; inflammatory disorder;
KW nervous system disorder; sexual dysfunction; pain; mood disorder;
KW hypertension; psychotic disorder; neurological disorder; dyskinesia;
KW viral infection; human immunodeficiency virus; HIV; non-viral infection;
KW ocular disease; cytostatic; enzyme.

OS Homo sapiens.

XX WO200200860-A2.

XX 03-JAN-2002.

XX 26-JUN-2001; 2001WO-US20171.

XX 26-JUN-2000; 2000US-214047P.

XX (SUGEN) SUGEN INC.

XX Florman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;

XX Chafydzak G;

XX

WPI; 2002-139913/18.
N-PSDB; ABK31757.
Nucleic acids encoding novel human proteases, useful for useful for treating diseases and disorders such as cancers, immune-related diseases and disorders, cardiovascular disease (e.g. restenosis) and inflammatory disorders -
Claim 6; Fig 2F-G; 313pp; English.
The present invention relates to the isolation of novel human proteases, and the nucleic acids encoding them. The sequences of the invention are useful for treating diseases and disorders such as cancers (e.g. breast, colon, lung), immune-related diseases and disorders (e.g. inflammatory diseases and asthma), cardiovascular diseases (e.g. restenosis and coronary thrombosis), brain or neuronal-associated diseases, metabolic disorders (e.g. diabetes, obesity), inflammatory disorders (e.g. rheumatoid arthritis and psoriasis), central or peripheral nervous system diseases, migraines, pain, sexual dysfunction, mood disorders, attention disorders, cognitive disorders, hypotension, hypertension, psychotic disorders, neurological disorders (e.g. Alzheimer's disease, Parkinson's disease) and dyskinesias. The nucleic acids and polypeptides are also useful for treating viral infections caused by human immunodeficiency virus (HIV), and non-viral infections such as ocular disease (e.g. glaucoma) and macular degeneration. AAU82702-AAU82760 represent the novel human proteases of the invention.

Query Match 100.0%; Score 8563; DB 23; Length 1604;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MGAKESRIGLSVEERARRVTVDELKRLKDAFRTCTGLSYMGQCHCFIREVLGDGVPKV 60
1 MGAKESRIGLSVEERARRVTVDELKRLKDAFRTCTGLSYMGQCHCFIREVLGDGVPKV 60
61 AEVYCSFGTSGKLHFNLLVGLVLTTRGKDEKAKYIFSLFSSBSGNVIREEMRL 120
61 AEVYCSFGTSGKLHFNLLVGLVLTTRGKDEKAKYIFSLFSSBSGNVIREEMRL 120
121 HVVDGKVPDLTRKCFSEGEKVKYEFKRNMLFNKDAFTFSRMLLSGGVYVTLTDDSDTPT 180
121 HVVDGKVPDLTRKCFSEGEKVKYEFKRNMLFNKDAFTFSRMLLSGGVYVTLTDDSDTPT 180
181 FYOTLAGVTHLESDDIIDLKRYLLKKAQSGRTGRFDETFGLVSPPIRPSLSEGLFNAP 240
181 FYOTLAGVTHLESDDIIDLKRYLLKKAQSGRTGRFDETFGLVSPPIRPSLSEGLFNAP 240
241 DENRDNDHDFKEISCGLSACCRGFLAERQKFCPKVDFDVRDGVLSRVELRDMVVALLEW 300
241 DENRDNDHDFKEISCGLSACCRGFLAERQKFCPKVDFDVRDGVLSRVELRDMVVALLEW 300
301 KDNRTDDIPELHNDLSIDIVEGINANDHTTQMGHLTIEDYQINSKYNLANEPLNLLFQVC 360
301 KDNRTDDIPELHNDLSIDIVEGINANDHTTQMGHLTIEDYQINSKYNLANEPLNLLFQVC 360
361 HIVLGLRAPATPEEGQIIRGWLERSRYGLQAGHNWFIISMQWQWQKEYVYKYPV 420
361 HIVLGLRAPATPEEGQIIRGWLERSRYGLQAGHNWFIISMQWQWQKEYVYKYPV 420
421 EPSVLNGGKYSFGTAAPHEQVEDRIGSSLSYVNTTEKFSNISTASEASETAGSGFL 480
421 EPSVLNGGKYSFGTAAPHEQVEDRIGSSLSYVNTTEKFSNISTASEASETAGSGFL 480
481 YSATPGADVCFARQHTSDNNQCLLGANGNILLNLPKPGALDNOPLVTPQPKATSL 540
481 YSATPGADVCFARQHTSDNNQCLLGANGNILLNLPKPGALDNOPLVTPQPKATSL 540
541 TLEGGRUKRTPLQIHGRDYEMVPEPVWRALYHWYGANLALPRVINKSKTDIPELELFFR 600
541 TLEGGRUKRTPLQIHGRDYEMVPEPVWRALYHWYGANLALPRVINKSKTDIPELELFFR 600

RESULT 2
ABG32160
ID ABG32160 standard; Protein; 1317 AA.

601 YLLFLRQOPATRTQQSNIWVWNGVPSNPAPLKRVLAYTCGFSRMOTIKIHEIYLSQRLR 660
601 YLLFLRQOPATRTQQSNIWVWNGVPSNPAPLKRVLAYTCGFSRMOTIKIHEIYLSQRLR 660
661 IKEDVRLWLYNSENVLTLLDDDEHKLKLYIKQDEQHLVIEVRNKMDSWPEMSFTANS 720
661 IKEDVRLWLYNSENVLTLLDDDEHKLKLYIKQDEQHLVIEVRNKMDSWPEMSFTANS 720
721 KIDRHKVPTEKGTATLSNLGNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRTNIGMK 780
721 KIDRHKVPTEKGTATLSNLGNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRTNIGMK 780
781 GHMAKYGDIVQELWLSGTQKNVAPLKLRTWIAKYAPRNGFQOQDSQELLAFLLDGLHED 840
781 GHMAKYGDIVQELWLSGTQKNVAPLKLRTWIAKYAPRNGFQOQDSQELLAFLLDGLHED 840
841 LNRVHEKPYVELKSDGRPDWEAAEANDNHLRNRISIVVDLPHGOLRSOVKCKTCHIS 900
841 LNRVHEKPYVELKSDGRPDWEAAEANDNHLRNRISIVVDLPHGOLRSOVKCKTCHIS 900
901 VRFPFNLPLPMDSYMHEITVTKLDTTPRVYGLRLNMDEKVTGLKKQSLDGLN 960
901 VRFPFNLPLPMDSYMHEITVTKLDTTPRVYGLRLNMDEKVTGLKKQSLDGLN 960
961 SEQILLAEVHGSNIKAPFQDNQKVRLSVSGFLCAFELPVPVSPISASSPTQTFSSPST 1020
961 SEQILLAEVHGSNIKAPFQDNQKVRLSVSGFLCAFELPVPVSPISASSPTQTFSSPST 1020
1021 NEMFTLTNGDLPRIPIPNMGMPNTVPCGTEKNTNGMVNGHMPSLPDSPTFYIAVH 1080
1021 NEMFTLTNGDLPRIPIPNMGMPNTVPCGTEKNTNGMVNGHMPSLPDSPTFYIAVH 1080
1081 RKMRTLYFLSSQKNRPSLFGMPLIVPCVTHTRKLDYDAVMIQVSRLASPLPQPEASN 1140
1081 RKMRTLYFLSSQKNRPSLFGMPLIVPCVTHTRKLDYDAVMIQVSRLASPLPQPEASN 1140
1141 HAQCDSDSMGYQYPTLTVVQKDGNSCAWCPWYRFRGCKIDCGEDRAFIAGNAYIAVDWD 1200
1141 HAQCDSDSMGYQYPTLTVVQKDGNSCAWCPWYRFRGCKIDCGEDRAFIAGNAYIAVDWD 1200
1201 PTALHLRYQTSQERVWDEHSEVQSRRAQAEPLNLSCLRAFTSEBELGENEMYCSKCK 1260
1201 PTALHLRYQTSQERVWDEHSEVQSRRAQAEPLNLSCLRAFTSEBELGENEMYCSKCK 1260
1261 THCLATKKLDLRLPPLIILHLKRFQVNGRWIKSQKIVKFPRESFDPSPAFIVPRDPALC 1320
1261 THCLATKKLDLRLPPLIILHLKRFQVNGRWIKSQKIVKFPRESFDPSPAFIVPRDPALC 1320
1321 QHKPLTPQGDLSPPRILAREVKVKAQSSAGBEDVLLSKSPSSLSANIISPKGSPSS 1380
1321 QHKPLTPQGDLSPPRILAREVKVKAQSSAGBEDVLLSKSPSSLSANIISPKGSPSS 1380
1381 RKSTCSPSSKNSSPSSPRTLGSKGRRLRLPOLGSKNKLSSKENLDASKENGAGQICE 1440
1381 RKSTCSPSSKNSSPSSPRTLGSKGRRLRLPOLGSKNKLSSKENLDASKENGAGQICE 1440
1441 LADALSRGHVGGSQPELVTPQDHEVALANGFYIEHACGNGYSGNGOLGNSBEDTDDQ 1500
1441 LADALSRGHVGGSQPELVTPQDHEVALANGFYIEHACGNGYSGNGOLGNSBEDTDDQ 1500
1501 REDTRIPIYNLYAISCHSGILGGGHVTVYAKPNCKWYCVNDSSCKELHPDEIDTDSAY 1560
1501 REDTRIPIYNLYAISCHSGILGGGHVTVYAKPNCKWYCVNDSSCKELHPDEIDTDSAY 1560
1561 ILFYEQQGIDYVAQFLPKTDGKQVADTSSMDEDESDYKCYVLQ 1604
1561 ILFYEQQGIDYVAQFLPKTDGKQVADTSSMDEDESDYKCYVLQ 1604

XX	ABG32160;	Best Local Similarity 99.9%; Pred. No. 0;			
AC	29-NOV-2002 (first entry)	Matches 1274; Conservative 1; Mismatches 0; Indels 0; Gaps 0;			
DT	Novel human ubiquitin carboxy-terminal hydrolase 80091.				
DE	Ubiquitin carboxy-terminal hydrolase; haematopoietic disorder; obesity;				
XX	anaemia; drug-induced anaemia; haemolytic anaemia; erythrocytosis;				
KW	chronic renal failure; erythrocytosis; cancer; neutropenia; anorexia;				
KW	granulocytopenia; AIDS; Acquired immunodeficiency syndrome; diabetes;				
KW	myelodysplastic syndrome; cell proliferation disorder; poliomyelitis;				
KW	cell differentiation disorder; neurological disorder; brain disorder;				
KW	cerebrovascular disease; acute meningitis; intracranial haemorrhage;				
KW	Alzheimer's disease; Parkinson's disease; Huntington's disease; tumour;				
KW	metabolic disorder; cachexia; lipid disorder; angiotensin disorder;				
KW	psoriasis; wound healing; rheumatoid arthritis; Kawasaki syndrome;				
KW	endothelial cell disorder; gene therapy; human.				
XX	Homo sapiens.				
OS	WO200263031-A2.				
PN	15-AUG-2002.				
XX	07-FEB-2002; 2002WO-US03793.				
PD	07-FEB-2001; 2001US-267054P.				
XX	(MILL-) MILLENNIUM PHARM INC.				
PA	Meyers RE;				
PI	WPI; 2002-627564/67.				
XX	N-PSDB; ABK90784.				
DR	Novel isolated ubiquitin carboxy-terminal hydrolase family polypeptide,				
XX	designated 80091 polypeptides, useful for detecting modulator compounds				
PT	for treating e.g. hematopoietic disorders, cancers and				
PT	granulocytopenias -				
XX	Claim 5; Page 112-113; 123pp; English.				
PS	The invention describes an isolated human ubiquitin carboxy-terminal				
XX	hydrolase family polypeptide, designated 80091 polypeptide (I). (I) is				
CC	useful as reagents or targets in assays applicable to treatment and				
CC	diagnosis of 80091-mediated or -related disorders. Antibodies are useful				
CC	for inhibiting the proliferation or inducing the killing of an 80091				
CC	expressing cell, for treating haematopoietic disorders e.g. anaemia such				
CC	as drug-induced anaemia, haemolytic anaemia, aberrant erythrocytosis,				
CC	anaemia of chronic disease such as chronic renal failure, erythrocytosis				
CC	and/or cancer. Antibodies are also useful for treating or preventing				
CC	neutropenia, granulocytopenia, AIDS, congenital and cyclic neutropenia,				
CC	and myelodysplastic syndromes. 80091 molecules are useful for treating				
CC	and diagnosing haematopoietic disorders, cellular proliferative and/or				
CC	differentiative disorders, neurological or brain disorders,				
CC	cerebrovascular diseases (e.g. acute meningitis, intracranial				
CC	haemorrhage, Alzheimer's disease, Parkinson's disease, Huntington's				
CC	disease and poliomyelitis), metabolic disorders (e.g. obesity, anorexia,				
CC	cachexia, diabetes and lipid disorders), angiogenic disorders (e.g.				
CC	psoriasis, wound healing and complications in diseases such as diabetes				
CC	and rheumatoid arthritis) and endothelial cell disorders (e.g. tumours				
CC	and Kawasaki syndrome). The polynucleotide encoding (I) is useful in gene				
CC	therapy, to express (I), to detect 80091 mRNA or a genetic alteration in				
CC	a 80091 gene, to modulate 80091 activity, in chromosome mapping, to				
CC	identify an individual from a minute biological sample (tissue typing),				
CC	and to aid in forensic identification of the biological sample. This				
CC	is the amino acid sequence of the novel human ubiquitin carboxy-terminal				
CC	hydrolase 80091.				
XX	Sequence 1317 AA;				
SQ	Query Match 79.7%; Score 6826; DB 23; Length 1317;				

1063 SAGEEDVLSKSPSSLSANIISPKGSPSSSRKSGTSCPPSKNSPNSPRTLGRSGRL 1122
1410 RLPQIGSKNKLSSSKENLDASKENGAGQICELADALSRGHVLGGSPQLVTPQDHEVALA 1469
1123 RLPQIGSKNKLSSSKENLDASKENGAGQICELADALSRGHVLGGSPQLVTPQDHEVALA 1182
1470 NGFLYEHEACGNGYNGQICNHSSEDDTDQREDTRIKPIYNLYAISCHSGLGGGHVY 1529
1183 NGFLYEHEACGNGYNGQICNHSSEDDTDQREDTRIKPIYNLYAISCHSGLGGGHVY 1242
1530 YAKNPKNKYCYNDSSCKELHDEIDTDSAYILFYEQQIDYVAQLPXTDGKQMDTSSM 1589
1243 YAKNPKNKYCYNDSSCKELHDEIDTDSAYILFYEQQIDYVAQLPXTDGKQMDTSSM 1302
1590 DEDFESDYKCYVLQ 1604
1303 DEDFESDYKCYVLQ 1317

RESULT 3
AAU75578
AAU75578 standard; Protein; 1274 AA.
AAU75578;
08-MAY-2002 (first entry)
Human ubiquitin specific protease 10 (USP10).
Ubiquitin specific protease 10; USP10; human; testis.
Homo sapiens.
CN1319662-A.
31-OCT-2001.
11-APR-2001; 2001CN-0113514.
11-APR-2001; 2001CN-0113514.
(UYNA-) UNIV NANJING MEDICAL.
Sha J, Zhou Z, Li J;
WPI; 2002-123068/17.
N-PSDB; ABK14799.
Human testicular ubiquitin specific proteinase 10 gene coding protein -
Claim 1; Page 1(Claims); 7pp; Chinese.
XX
The invention relates to a human testis ubiquitin specific protease 10,
(USP10) gene, and protein. The USP10 gene can be used to prepare a fusion
protein, which can be used to prepare mono-/polyclonal antibody. The gene
can be also be used for preparing testis specific function gene
expression chip, in gene therapy for curing USP10-related diseases. The
present sequence represents the amino acid sequence of human USP10.
XX
Sequence 1274 AA;
Query Match 79.5%; Score 6807; DB 23; Length 1274;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1272; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 331 MGLTLEDYQIWSVKNVLANEFANLIFQVCHIVGLRPAATPEEGQIIRGWLERSRYGL 390
DB 1 MGLTLEDYQIWSVKNVLANEFANLIFQVCHIVGLRPAATPEEGQIIRGWLERSRYGL 60
2Y 391 OAGHNWFIISMQWQWQWKEYKYKDANPVLPSPSSVLNGKYSFGTAHPMEQVEDRIGSS 450
DB 61 OAGHNWFIISMQWQWQWKEYKYKDANPVLPSPSSVLNGKYSFGTAHPMEQVEDRIGSS 120

QY 451 LSVYNTTEKFSNISTASEASETAGSGFLYSATPGADVCFARQHNSTNNNQCLLGANG 510
DB 121 LSVYNTTEKFSNISTASEASETAGSGFLYSATPGADVCFARQHNSTNNNQCLLGANG 180
QY 511 NIILHLNPOKPGAIHQPLVTOEPVKATSLTLEGGRLKRTPOLIHGRDYEMVPEPVWAL 570
DB 181 NIILHLNPOKPGAIHQPLVTOEPVKATSLTLEGGRLKRTPOLIHGRDYEMVPEPVWAL 240
QY 571 YHWYGANLALPREVIKNSKTDIPELFPYRILFRQQPATRTOQSNIWVNMGNVSPNA 630
DB 241 YHWYGANLALPREVIKNSKTDIPELFPYRILFRQQPATRTOQSNIWVNMGNVSPNA 300
QY 631 PLKRVLAYTCFSRMTIKIIEHYLSQRIRIKEEDMRWLXNSENYLTLLDDEHKLLEYL 690
DB 301 PLKRVLAYTCFSRMTIKIIEHYLSQRIRIKEEDMRWLXNSENYLTLLDDEHKLLEYL 360
QY 691 KIODEQLVIEVRNKMSPPEMSFIANSKIDRHKVPTKEGATGLSNLGNTCFMNSSIQ 750
DB 361 KIODEQLVIEVRNKMSPPEMSFIANSKIDRHKVPTKEGATGLSNLGNTCFMNSSIQ 420
QY 751 CVSNTOPLTOYFISGRHLYELNETNPIGMKGHWAKCYGDLVOELMSGTKQNVAPLKLRT 810
DB 421 CVSNTOPLTOYFISGRHLYELNETNPIGMKGHWAKCYGDLVOELMSGTKQNVAPLKLRT 480
QY 811 IAKYAPRFNGFQOQDSQELLAFLLDGLHEDLNARVHEKPYVELKSDGRPDWEAAEAWN 870
DB 481 IAKYAPRFNGFQOQDSQELLAFLLDGLHEDLNARVHEKPYVELKSDGRPDWEAAEAWN 540
QY 871 HLRRNRSIIVDLFPHGQIRGQVCKTCHISVAPDPNFSLPLPMDSYMHELTIVIKLDG 930
DB 541 HLRRNRSIIVDLFPHGQIRGQVCKTCHISVAPDPNFSLPLPMDSYMHELTIVIKLDG 600
QY 931 TTPVRYGLRLNMDKVTGLKQSLDCLNSEQILLAEVHGSNIKNFPQDNQKRLSVSG 990
DB 601 TTPVRYGLRLNMDKVTGLKQSLDCLNSEQILLAEVHGSNIKNFPQDNQKRLSVSG 660
QY 991 FLCAFEIPVPSISASPTQTDFFSSPSTNEMFTLTNGDLPRPIFINGMPTNVVPCG 1050
DB 661 FLCAFEIPVPSISASPTQTDFFSSPSTNEMFTLTNGDLPRPIFINGMPTNVVPCG 720
QY 1051 TEKNFTNGMVGNGHMPSLPDSPTFGYIIAHRKMRTELFLSSQKRPFLFGMLIVPCT 1110
DB 721 TEKNFTNGMVGNGHMPSLPDSPTFGYIIAHRKMRTELFLSSQKRPFLFGMLIVPCT 780
QY 1111 VHTRKKDLYDAVMIQVSRLASPLPQEAASNAHQDSDSMGYQYPTFLRVVQKGNCAWC 1170
DB 781 VHTRKKDLYDAVMIQVSRLASPLPQEAASNAHQDSDSMGYQYPTFLRVVQKGNCAWC 840
QY 1171 PWYRRCGCKIDCGEDRAFIQNAIYIADWDPTALHLYOTSQSERVVDEHESVQSRRAQ 1230
DB 841 PWYRRCGCKIDCGEDRAFIQNAIYIADWDPTALHLYOTSQSERVVDEHESVQSRRAQ 900
QY 1231 EPINLDSCLRAFTSEELGENEMYCYCKKTHCLATKCLDLMLRPLPILIIHLKRFVNG 1290
DB 901 EPINLDSCLRAFTSEELGENEMYCYCKKTHCLATKCLDLMLRPLPILIIHLKRFVNG 960
QY 1291 RWIKSQIKVFPRESPPSAFLVPRPALCOHPLTPQGDSELPSEPLAREVKVKVAQSS 1350
DB 961 RWIKSQIKVFPRESPPSAFLVPRPALCOHPLTPQGDSELPSEPLAREVKVKVAQSS 1020
QY 1351 AGEEDVLLSKSPSSLSANIISPKGSPSSSRKSGTSCPPSKNSPNSPRTLGRSGRL 1410
DB 1021 AGEEDVLLSKSPSSLSANIISPKGSPSSSRKSGTSCPPSKNSPNSPRTLGRSGRL 1080
QY 1411 LPQIGSKNKLSSSKENLDASKENGAGQICELADALSRGHVLGGSPQLVTPQDHEVALA 1470
DB 1081 LPQIGSKNKLSSSKENLDASKENGAGQICELADALSRGHVLGGSPQLVTPQDHEVALA 1140
QY 1471 GFLYEHEACGNGYNGQICNHSSEDDTDQREDTRIKPIYNLYAISCHSGLGGGHVY 1530
DB 1141 GFLYEHEACGNGYNGQICNHSSEDDTDQREDTRIKPIYNLYAISCHSGLGGGHVY 1200
QY 1531 AKNPCKWYCYNDSSCKELHDEIDTDSAYILFYEQQIDYVAQLPXTDGKQMDTSSM 1590

Db 1201 AXNPNCKWYCYNDSSCKELHPDEIDTBSAYILFVEQOQIDYALPXTDGGKQVADTSSMD 1260
Qy 1591 EDFESDYKKYCVLQ 1604
Db 1261 EDFESDYKKYCVLQ 1274
RESULT 4
ABU12111
ID ABU12111 standard; Protein; 1059 AA.
XX
AC ABU12111;
XX
DT 17-FEB-2003 (first entry)
XX
DE Human protein modification and maintenance molecule (PMOD) #8.
XX
KW Human; protein modification and maintenance molecule; PMOD;
KW gastrointestinal disorder; anorexia; nausea; cardiovascular disorder;
KW angina pectoris; myocardial infarction; autoimmune disorder; anaemia;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS; gout;
KW developmental disorder; epilepsy; epithelial disorder; vitiligo; eczema;
KW neurological disorder; Huntington's disease; Parkinson's disease;
KW reproductive disorder; infertility; impotence; Inceyte 7485451CD1.
XX
OS Homo sapiens.
XX
PN WO200281636-A2.
XX
PD 17-OCT-2002.
XX
PF 05-APR-2002; 2002WO-US10812.
XX
PR 05-APR-2001; 2001US-282282P.
PR 13-APR-2001; 2001US-283782P.
PR 18-APR-2001; 2001US-284823P.
PR 27-APR-2001; 2001US-287264P.
PR 04-MAY-2001; 2001US-288662P.
PR 11-MAY-2001; 2001US-290383P.
PR 15-JUN-2001; 2001US-298348P.
PR 25-JAN-2002; 2002US-351928P.
PR 25-FEB-2002; 2002US-359903P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Deleage AM, Swarnakar A, Hafalia AJA, Duggan BM;
PI Warren BA, Emerling BM, Arvizu CS, Honchell CD, Kallick DA;
PI Lu DAM, Lee EA, Yue H, Forsythe J, Ramkumar J, Griffin JA;
PI Li JX, Thangavelu K, Baughn MR, Yao MG, Sanjanwala MM, Wallia NK;
PI Burford N, Lal PG, Becha SD, Lee SY, Elliott VS, Luo W, Lu Y;
PI Wang YE;
XX
DR WPI; 2003-058519/05.
DR N-PSDB; ABX56506.
XX
PT Novel isolated human protein modification and maintenance molecules
PT useful for diagnosing, treating, preventing Alzheimer's disease,
PT epilepsy, gout, impotence, anaemia, angina pectoris, vitiligo, anorexia,
PT nausea
XX
PS Claim 1; Page 147-149; 184pp; English.
XX
CC The invention relates to human protein modification and maintenance
CC molecules (PMOD) and to polynucleotides encoding them. The polypeptides
CC of gastrointestinal disorders (e.g. anorexia, nausea, treatment and prevention
CC of disorders (e.g. angina pectoris, myocardial infarction),
CC autoimmune/inflammatory disorders (e.g. acquired immunodeficiency
CC syndrome (AIDS), gout), developmental disorders (e.g. anaemia, epilepsy),
CC epithelial disorders (e.g. vitiligo, eczema), neurological disorders
CC (e.g. Huntington's disease, Parkinson's disease) and reproductive
CC disorders (e.g. infertility, impotence). The polypeptides may be used in

CC assays for detecting the presence of the associated disorders. The
CC polynucleotides are useful for detecting upstream sequences such as
CC promoters and regulatory elements, for creating knock out or knock in
CC humanised animals or transgenic animals to model human diseases and in
CC somatic or germline gene therapy for treating the disorders. Sequences
CC ABU12104-ABU12120 represent human protein modification and maintenance
XX molecules of the invention.
SQ Sequence 1059 AA;
Query Match 65.1%; Score 5570.5; DB 24; Length 1059;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1044; Conservative 0; Mismatches 0; Indels 15; Gaps 2;
Qy 561 MYPEPWALYHWYGANLALPRVILKNSKTIDPELELPYLLFLRQOPATRTQOSNIWV 620
Db 1 MYPEPWALYHWYGANLALPRVILKNSKTIDPELELPYLLFLRQOPATRTQOSNIWV 60
Qy 621 NM-----GNVPSNAPLKRVLAYTGCFSRMOTIKEIHEYLSQRLRIKEEDM 666
Db 61 NMGMWSLRFPQHLPRGNVPSNAPLKRVLAYTGCFSRMOTIKEIHEYLSQRLRIKEEDM 120
Qy 667 RLWLYNSENLYLLDDEHKLKLYLKIODEHLYVIVRNKMSWPEMSFIANSKIDRHK 726
Db 121 RLWLYNSENLYLLDDEHKLKLYLKIODEHLYVIVRNKMSWPEMSFIANSKIDRHK 180
Qy 727 VPEKGATGSLNLTGNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRPINIGMKGMAC 786
Db 181 VPEKGATGSLNLTGNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRPINIGMKGMAC 240
Qy 787 YGDLVQELWSGTQKNVAPLKLWTTAKYAPRNGQQOQSDELLAFLDLGLHEDLNRVHE 846
Db 241 YGDLVQELWSGTQKNVAPLKLWTTAKYAPRNGQQOQSDELLAFLDLGLHEDLNRVHE 300
Qy 847 KPYVELKDSGDRPDWEVAEAWNDHLRNRSLVVDLPHGQLRSQVCKCTCGHISVRFDPF 906
Db 301 KPYVELKDSGDRPDWEVAEAWNDHLRNRSLVVDLPHGQLRSQVCKCTCGHISVRFDPF 360
Qy 907 NFLSLPLMDSYMHLEITVIKLDGTPVRYGURLNNDKTYTGLKKQLSDLCGLNSEQILL 966
Db 361 NFLSLPLMDSYMHLEITVIKLDGTPVRYGURLNNDKTYTGLKKQLSDLCGLNSEQILL 420
Qy 967 AEVHGSNIKNFPQDNQKRLSVSGFLCAFEIIPVPSPIASSTPTQDFSSSSTNEMFTL 1026
Db 421 AEVHGSNIKNFPQDNQKRLSVSGFLCAFEIIPVPSPIASSTPTQDFSSSSTNEMFTL 480
Qy 1027 TTNGDLPRFIFIPNGMNPNTVPCGTEKNTNGMNGHMPSPDPSPTGYIIAVHKKMRT 1086
Db 481 TTNGDLPRFIFIPNGMNPNTVPCGTEKNTNGMNGHMPSPDPSPTGYIIAVHKKMRT 540
Qy 1087 ELYFLSSQKRPFLGMPPLIVPCTVHTRKQDLYDAVWQVSLASPLPPQEASNAHQDCC 1146
Db 541 ELYFLSSQKRPFLGMPPLIVPCTVHTRKQDLYDAVWQVSLASPLPPQEASNAHQDCC 600
Qy 1147 DSMGYQYPTLRAVQKNSCAWCWYRFCRCCKIDCGEDRAFIQNAIYAVDWDPTLHL 1206
Db 601 DSMGYQYPTLRAVQKNSCAWCWYRFCRCCKIDCGEDRAFIQNAIYAVDWDPTLHL 660
Qy 1207 RYQTSQERVDEHSEVQSRRAQAEPIINLDSCLRAFTSEELGENEMYCSCKKTHCLAT 1266
Db 661 RYQTSQERVDEHSEVQSRRAQAEPIINLDSCLRAFTSEELGENEMYCSCKKTHCLAT 720
Qy 1267 KKLDELWLPILIIHLKRPQVNGEWIKSOKIVKPFRESFDSALVPRDRPALCOHKPLT 1326
Db 721 KKLDELWLPILIIHLKRPQVNGEWIKSOKIVKPFRESFDSALVPRDRPALCOHKPLT 780
Qy 1327 PQGDELSEPRILAREVKKYVDAQSSAGEEDVLLSKSPSSLSANIISSPKGPSSSSRKSGTS 1386
Db 781 PQGDELSEPRILAREVKKYVDAQSSAGEEDVLLSKSPSSLSANIISSPKGPSSSSRKSGTS 840
Qy 1387 CPSSKNSSPNSSPRLTGRSKGRRLRPLQIGSKNKLSSSKENLDASKENGAGQICELADALS 1446
Db 841 CPSSKNSSPNSSPRLTGRSKGRRLRPLQIGSKNKLSSSKENLDASKENGAGQICELADALS 900

QY 1447 RGHVLG-GSQPELVTPQDHEVALANGFLYHEACGNGYNGQIGNHSEEDSTDQREDTR 1505
 DB 901 RGHVLGVSQPELVTPQDHEVALANGFLYHEACGNGYNGQIGNHSEEDSTDQREDTR 960
 QY 1506 IKPIYNLYAISCHSGILGGHYVTYAKNPCKWYCYNDSSCKELHPDEIDTDSAYILFYE 1565
 DB 961 IKPIYNLYAISCHSGILGGHYVTYAKNPCKWYCYNDSSCKELHPDEIDTDSAYILFYE 1020
 QY 1566 QQGIDYAQFLPKTDGKWADTSSMDEDFESDYKICYVLQ 1604
 DB 1021 QQGIDYAQFLPKTDGKWADTSSMDEDFESDYKICYVLQ 1059

RESULT 5

AAM40477

ID AAM40477 standard; Protein; 1127 AA.

XX AC AAM40477;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 5408.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Sky-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX XX 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX XX (HYSE-) HYSEQ INC.

XX PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;
 XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX XX WPI; 2001-442253/47.

XX DR N-PSDB; AAI59633.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders
 XX PT such as central nervous system injuries -

XX PS Example 2; SEQ ID NO 5408; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Sky-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX SQ Sequence 1127 AA;

Query Match 56.3%; Score 4822; DB 22; Length 1127;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 901; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 701 EVRNKDMSPREMSFIANSKKIDRHKVPTEKGATGSLNLTGTCFNMSSIQCVSNTPLTQ 760

DB 220 EVRNKDMSPREMSFIANSKKIDRHKVPTEKGATGSLNLTGTCFNMSSIQCVSNTPLTQ 279

QY 761 YFISGRHLYELNRTNPIGMKGHMAKCYGDLVQELWSGTQKNVAPLKLRTWTIAKYPFNG 820

DB 280 YFISGRHLYELNRTNPIGMKGHMAKCYGDLVQELWSGTQKNVAPLKLRTWTIAKYPFNG 339

QY 821 FOQODSQELLAFLLDGLHEDLNVRHEKPYVELKDSGDPDWEVAAEADNHLNRNRSIV 880

DB 340 FOQODSQELLAFLLDGLHEDLNVRHEKPYVELKDSGDPDWEVAAEADNHLNRNRSIV 399

QY 881 DLFHQQLRSQVKCKTCGHSIVRFPFNFSLPLPMDSYMHLEITVIKLDGTPVRYGLRL 940

DB 400 DLFHQQLRSQVKCKTCGHSIVRFPFNFSLPLPMDSYMHLEITVIKLDGTPVRYGLRL 459

QY 941 NMDEKYGKQLSDLCGLNSEQILLAEVHGSNKNPQDNQKVRLSVSGFLCAFEIPEVP 1000

DB 460 NMDEKYGKQLSDLCGLNSEQILLAEVHGSNKNPQDNQKVRLSVSGFLCAFEIPEVP 519

QY 1001 VSPISASSPTQTDSSSPSTNEMFTLTNGDLPRPIIPNGMNPNTVPCGTEKNFTNGMV 1060

DB 520 VSPISASSPTQTDSSSPSTNEMFTLTNGDLPRPIIPNGMNPNTVPCGTEKNFTNGMV 579

QY 1061 NGHMPSLPDSPTGYIIAIVHRKMRTELYFLSSQKNRPSLFGMPLIIVPCTVHTKKOLDYD 1120

DB 580 NGHMPSLPDSPTGYIIAIVHRKMRTELYFLSSQKNRPSLFGMPLIIVPCTVHTKKOLDYD 639

QY 1121 AVWIOVSRSLASPLPQEAASNAQDCDDSMGYQYPTLAVVQKDGNSCAWCWPYFCRGCK 1180

DB 640 AVWIOVSRSLASPLPQEAASNAQDCDDSMGYQYPTLAVVQKDGNSCAWCWPYFCRGCK 699

QY 1181 IDCGEDRAFIIGNAYIAVDWPTALHLRYQTSQERVVDEHESVEQSRAQAQVEPINLDSCLR 1240

DB 700 IDCGEDRAFIIGNAYIAVDWPTALHLRYQTSQERVVDEHESVEQSRAQAQVEPINLDSCLR 759

QY 1241 APTSEELGENEMYCSKCKTHCLATKCLDLWRPLPILIIHLKRFQFVNGRWIKSQIKV 1300

DB 760 APTSEELGENEMYCSKCKTHCLATKCLDLWRPLPILIIHLKRFQFVNGRWIKSQIKV 819

QY 1301 PPRSFDFPSAFIVPRDPALCOHKPLTPQGDLSLPRILAREVKVDAQSSAGEDVLLSK 1360

DB 820 PPRSFDFPSAFIVPRDPALCOHKPLTPQGDLSLPRILAREVKVDAQSSAGEDVLLSK 879

QY 1361 SPSSLSANIISSPKGSPSSSRKSGTSCPSKNSPNSSPRTLGRSKORLRLPQIGSKNKL 1420

DB 880 SPSSLSANIISSPKGSPSSSRKSGTSCPSKNSPNSSPRTLGRSKORLRLPQIGSKNKL 939

QY 1421 SSSKENLDASKENGAGQICELADALSGRHVGGSOPELVTPQDHEVALANGFLYEHEA-- 1478

DB 940 SSSKENLDASKENGAGQICELADALSGRHVGGSOPELVTPQDHEVALANGFLYEHEA-- 999

QY 1479 --CGNGYNGQIGNHSEEDSTDQREDTRIKPIYNLYAISCHSGILGGHYVTYAKNPNC 1536

DB 1000 NGCGNGYNGQIGNHSEEDSTDQREDTRIKPIYNLYAISCHSGILGGHYVTYAKNPNC 1059

QY 1537 KYCYNDSSCKELHPDEIDTDSAYILFYEQIGIDYAOFLPKTDGKWADTSSMDEDFESD 1596

DB 1060 KYCYNDSSCKELHPDEIDTDSAYILFYEQIGIDYAOFLPKTDGKWADTSSMDEDFESD 1119

QY 1597 YKCYCVLQ 1604

DB 1120 YEKYCVLQ 1127
|:|||||
RESULT 6
AA040478
ID AAM40478 standard; Protein; 1127 AA.
AC AAM40478;
XX
DT 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 5409.
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
XX 09-JUL-2000; 2000US-0552317.
XX 19-JUL-2000; 2000US-0598042.
XX 03-AUG-2000; 2000US-0620312.
XX 14-SEP-2000; 2000US-0653450.
XX 19-OCT-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Dermanac RT;
XX WPI; 2001-442253/47.
XX N-PSDB; AA159634.
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 5409; 10078pp; English.
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AA42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 1127 AA;
XX
XX Query Match 56.3%; Score 4822; DB 22; Length 1127;
XX Best Local Similarity 99.2%; Pred. No. 0;
XX

Matches	901; Conservative	1; Mismatches	2; Indels	4; Gaps	1;
QY	701	EVNKMDSWPEEMSF	FIANSKIDRHKVPT	KEGATGLNGLNTCF	MNSSIQCVSNTQPLTQ 760
Db	220	EVNKMDSWPEEMSF	FIANSKIDRHKVPT	KEGATGLNGLNTCF	MNSSIQCVSNTQPLTQ 279
QY	761	YFISGRHLYELANT	PIGKGMKACYGDL	VOELWSGTQKVAPL	KLRWTIAKYAPRNG 820
Db	280	YFISGRHLYELANT	PIGKGMKACYGDL	VOELWSGTQKVAPL	KLRWTIAKYAPRNG 339
QY	821	FQODSQELLAFLL	DGLHEDLNVRHEK	PYVELKDSGRPD	WEAAEAWNDHLRRRSIVV 880
Db	340	FQODSQELLAFLL	DGLHEDLNVRHEK	PYVELKDSGRPD	WEAAEAWNDHLRRRSIVV 399
QY	881	DLPHGLRQVQKCT	CGHISVRFDPNF	LSLPMDSYMHLEIT	VIKLDGTTTPVRYGLR 940
Db	400	DLPHGLRQVQKCT	CGHISVRFDPNF	LSLPMDSYMHLEIT	VIKLDGTTTPVRYGLR 459
QY	941	NMDEKTYTGLKK	LSLDCGLNSEQI	LLAEVHGNIKNF	PCDNOKVRLSVSGFLCAFEI 1000
Db	460	NMDEKTYTGLKK	LSLDCGLNSEQI	LLAEVHGNIKNF	PCDNOKVRLSVSGFLCAFEI 519
QY	1001	VSPISASPTQTD	FSSPSTNEMFTL	TNGDLPRFIFIP	NGMPNTVVPCKTEKNTGMV 1060
Db	520	VSPISASPTQTD	FSSPSTNEMFTL	TNGDLPRFIFIP	NGMPNTVVPCKTEKNTGMV 579
QY	1061	NGHMPSLDPSPT	GYIIAHRKMMTE	LYFLSSQKNRPSL	FCMPLIVPCTVTRKKOLYD 1120
Db	580	NGHMPSLDPSPT	GYIIAHRKMMTE	LYFLSSQKNRPSL	FCMPLIVPCTVTRKKOLYD 639
QY	1121	AVMIQVSRLASP	LPQEAASNAQD	CCDSMGYQYPT	FLRVVQKDGNSCAWCPYR 1180
Db	640	AVMIQVSRLASP	LPQEAASNAQD	CCDSMGYQYPT	FLRVVQKDGNSCAWCPYR 699
QY	1181	IDCGEDRAFI	GNAYTAVDWDPT	ALHLRYOTSERV	VDHESVEQSRAQABINL 1240
Db	700	IDCGEDRAFI	GNAYTAVDWDPT	ALHLRYOTSERV	VDHESVEQSRAQABINL 759
QY	1241	AFTSEELGENE	MYCCKTKHCLAT	KKLDLWRLPPLI	HLIKRFQFVNGRWIKSQI 1300
Db	760	AFTSEELGENE	MYCCKTKHCLAT	KKLDLWRLPPLI	HLIKRFQFVNGRWIKSQI 819
QY	1301	FPRESFPSAFL	VPRDPALCQHKP	TPQGEUSEPRIL	AREVKVDAQSSAGEEDV 1360
Db	820	FPRESFPSAFL	VPRDPALCQHKP	TPQGEUSEPRIL	AREVKVDAQSSAGEEDV 879
QY	1361	SPSSLSANTI	ISSPKGSPSSRS	KSGTSCPSKNS	SPSSPRTLGRSGRLRL 1420
Db	880	SPSSLSANTI	ISSPKGSPSSRS	KSGTSCPSKNS	SPSSPRTLGRSGRLRL 939
QY	1421	SSSKENLDAS	KENGAGQICELAD	ALSRGHVGGSOPE	LVTPQDHEVALANGFLYE 1478
Db	940	SSSKENLDAS	KENGAGQICELAD	ALSRGHVGGSOPE	LVTPQDHEVALANGFLYE 999
QY	1479	--CGGYNSGQI	GNHSEEDTDQRE	DTRIKPIYNYL	YAIISCHGILGGHYTYA 1536
Db	1000	NGCGYNSGQI	GNHSEEDTDQRE	DTRIKPIYNYL	YAIISCHGILGGHYTYA 1059
QY	1537	KWCVNDSSCKE	LHPDEITDSDAY	ILFYEQQIGIDYA	QFLPKTDGKKMADTSM 1596
Db	1060	KWCVNDSSCKE	LHPDEITDSDAY	ILFYEQQIGIDYA	QFLPKTDGKKMADTSM 1119
QY	1597	YKKYCVLQ 1604			
Db	1120	YKKYCVLQ 1127			
RESULT 7					
AAM38692					
ID	AAM38692 standard; Protein; 1085 AA.				
XX					
AC	AAM38692;				
XX					

DT	22-OCT-2001	(first entry)	
XX	Human polypeptide SEQ ID NO 1837.		
XX	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
OS	Homo sapiens.		
XX	WO200153312-A1.		
PN	26-JUL-2001.		
XX	26-DEC-2000; 2000WO-US34263.		
XX	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	03-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX	(HYSE-) HYSE INC.		
PA	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
XX	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX	WPI; 2001-442253/47.		
DR	N-PSDB; AA157848.		
XX	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX	Example 3; SEQ ID NO 1837; 10078pp; English.		
XX	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AA138642-AA142213) with neotropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukaemias and		
CC	C.N.S disorders.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX	Sequence 1085 AA;		
XX	Query Match 53.1%; Score 4543; DB 22; Length 1085;		
XX	Best Local Similarity 94.1%; Pred. No. 0;		
XX	Matches 851; Conservative 22; Mismatches 31; Indels 0; Gaps 0;		
QY	701 EVRNKMSWPEEMSFANSSKIDRHKVPTKGTGLNLGNTCFMNSSIQCVSNTQPLTQ 760		
DB	182 EVHNKMSWPEEMSFANSSKIDRHKVPTKGTGLNLGNTCFMNSSIQCVSNTQPLTQ 241		
QY	761 YFISGRHLVELNRTNPTGMKGMAKCYGDLVQELWSGTQKVAFLKLRWTIAKAPFNG 820		
DB	242 YFISGRHLVELNRTNPTGMKGMAKCYGDLVQELWSGTQKVAFLKLRWTIAKAPFNG 301		
QY	821 FQQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEAAEAWDNHLNRNSIVV 880		
DB	302 FQQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEAAEAWDNHLNRNSIVV 361		
QY	881 DLPHGQLRSQVKCKTCGHISVRDPNPNFLSLPLPMSYHLEITVILKLDGTTVPVYGLRL 940		
DB	362 DLPHGQLRSQVKCKTCGHISVRDPNPNFLSLPLPMSYHLEITVILKLDGTTVPVYGLRL 421		
QY	941 NMDEKTYGLKKQSLDCLGNSQLLAEVHGSNIKPFQDNQKVRLSVSGFLCAFEIPVP 1000		
DB	422 NMDEKTYGLKKQSLDCLGNSQLLAEVHGSNIKPFQDNQKVRLSVSGFLCAFEIPVP 481		
QY	1001 VSPISASSPTQTPSSSPSTNEMFTLTNGDLPRPIPIKNGMNTVVPCTGTEKNTNGMV 1060		
DB	482 SSPISASSPTQIDFSSSPSTNGMFTLTNGDLPRPIPIKNGMNTVVPCTGTEKNTNGMV 541		
QY	1061 NGHMPSLPDSPTFTGYIIAVHRKWMRTLYFLSKQKRPISLFGMPLIIVPCTVHTKQKLDYD 1120		
DB	542 NGHMPSLPDSPTFTGYIIAVHRKWMRTLYFLSKQKRPISLFGMPLIIVPCTVHTKQKLDYD 601		
QY	1121 AVTIQVSRSLASPLPPQASNAHQDSDSMGYQYPTLTVVKQKGNCSAMCPWTRFCRGCK 1180		
DB	602 AVTIQVSRSLASPLPPQASNAHQDSDSMGYQYPTLTVVKQKGNCSAMCPWTRFCRGCK 661		
QY	1181 IDCGEDRAFTGNAYIAVDWPTALHLRYQTSQSRVVDHESVQSRRAAQAEPINLDSCLR 1240		
DB	662 IDCGEDRAFTGNAYIAVDWPTALHLRYQTSQSRVVDHESVQSRRAAQAEPINLDSCLR 721		
QY	1241 AFTSEEBELGENEMYCYCKKTHCLATKLDLWRLPPLIIHLKRFQFVNGRWINKSQKIVK 1300		
DB	722 AFTSEEBELGENEMYCYCKKTHCLATKLDLWRLPPLIIHLKRFQFVNDQWIKSQKIVR 781		
QY	1301 FPRSFDPSPAFVPRDPAALCOHKPLTPQGDSELPRLABEVKKVDAQSSAGDEEDVLSK 1360		
DB	782 FLRSFDPSPAFVPRDPAALCOHKPLTPQGDSELPRLABEVKKVDAQSSAGDEEDVLSK 841		
QY	1361 SPSSLSANISSPKGSPSSSRKSGTSCFSSKNSSPNSPRTLGSKGRLRLPQIGSKNKL 1420		
DB	842 SPSSLSANISSPKGSPSSSRKSGTSCFSSKNSSPNSPRTLGSKGRLRLPQIGSKNKP 901		
QY	1421 SSSKENLDASKENGAGQICELADALSRGHVLGSGQPELVTPQDHEVALANGFLYEACG 1480		
DB	902 SSSKENLDASKENGAGQICELADALSRGHVLGSGQPELVTPQDHEVALANGFLYEACG 961		
QY	1481 NGYNGQLGNHSESDTDDQREDTRIKPIYNLVAISCHSILGSGHYVTVAKPNCKWYC 1540		
DB	962 NGYNGQLGNHSESDTDDQREDTRIKPIYNLVAISCHSILGSGHYVTVAKPNCKWYC 1021		
QY	1541 YNDSCKELHPDEITDTSAYILFVEQGGIDYAPLPKTDGKKWADTSSMDEDESDYKKY 1600		
DB	1022 YNDSCKELHPDEITDTSAYILFVEQGGIDYAPLPKTDGKKWADTSSMDEDESDYKKY 1081		
QY	1601 CVLQ 1604		
DB	1082 SMLQ 1085		
XX	RESULT 8		
XX	AAW82396		
XX	ID AAW82396 standard; Protein; 1089 AA.		
XX	AC AAW82396;		
XX	DT 23-FEB-1999 (first entry)		
XX	XX Human UB protein #2.		
XX	DE Ubiquitin isopeptidase; UB; intracellular protein; degradation; therapy;		
KW	cell growth; diagnosis; tumour; proliferative disease.		
OS	Homo sapiens.		
XX	WO9848020-A2.		

XX PD 29-OCT-1998.
XX XX
PF 10-APR-1998; 98WO-IT00084.
XX XX
PR 18-APR-1997; 97IT-MI00905.
XX XX
PA (EUON-) IST EURO DI ONCOLOGIA SRL.
XX XX
PI Bosari S, Capra M, Di Fiore PP, Draetta GF, Goubin P,
PI Matoskova B, Naviglio S, Soncini C;
XX XX
DR WPI; 1998-609991/51.
XX XX
PS N-PSDB; AAV73472.
XX XX
PT New ubiquitin isopeptidases that control cell proliferation - useful
PT in the diagnosis and treatment of tumours and other proliferative
PT conditions
XX XX
PS Claim 1; Page 44-48; 75pp; English.
XX XX
SQ This sequence represents a novel human ubiquitin isopeptidase, UBP.
CC Ubiquitin isopeptidases are involved in the ubiquitin pathway which
CC controls the degradation of a large number of intracellular proteins,
CC many of which play key roles in controlling cell growth. Antibodies
CC raised against this protein may be used in the diagnosis of a tumour or
CC another proliferative disease or for therapeutic or diagnostic purposes.
XX XX
SQ Sequence 1089 AA;
Query Match 52.9%; Score 4531; DB 19; Length 1089;
Best Local Similarity 93.7%; Pred. No. 0;
Matches 851; Conservative 22; Mismatches 31; Indels 4; Gaps 1;
701 EVRNKMSWPEEESFIANSKIDRHKVTEKGTGLNLTGNTCFMNSSIOCVNTQPLTQ 760
182 EVRNKMSWPEEESFIANSKIDRHKVTEKGTGLNLTGNTCFMNSSIOCVNTQPLTQ 241
761 YFISGRHLYELNRTNPIMKGHMAKVGDIIVQELWSGTQRNAPLKLRTWTIAKAPRFG 820
242 YFISGRHLYELNRTNPIMKGHMAKVGDIIVQELWSGTQRNAPLKLRTWTIAKAPRFG 301
821 FQODSDELLAFLLDGLHEDLNVEHXPYVELKDSGRPDWEVAABAWNDHLRNRISIV 880
302 FQODSDELLAFLLDGLHEDLNVEHXPYVELKDSGRPDWEVAABAWNDHLRNRISIV 361
881 DLPHGLRSQVCKCTGHI SVRDPDFNFLSLPLPMDSYMHELEITVIKLDGTTVRYGLRL 940
362 DLPHGLRSQVCKCTGHI SVRDPDFNFLSLPLPMDSYMHELEITVIKLDGTTVRYGLRL 421
941 NMDEKTYGLKQLSDICGLNSEQILLAEVHGSNIKPPQDNQKRLSVSGFLCAFEIPVP 1000
422 NMDEKTYGLKQLSDICGLNSEQILLAEVHGSNIKPPQDNQKRLSVSGFLCAFEIPVP 481
1001 VSPISASSPTQDFSSSPSTNEMFTLTNGDLPRPIFINGMENTVVPCKTEKFTNGMV 1060
482 SPTISASSPTQDFSSSPSTNEMFTLTNGDLPRPIFINGMENTVVPCKTEKFTNGMV 541
1061 NGHMPSLPDPTGTYIAVHRKWRTELYFLSSQKRPRLFGLMPLIVPCTVHTRKDLVD 1120
542 NGHMPSLPDPTGTYIAVHRKWRTELYFLSSQKRPRLFGLMPLIVPCTVHTRKDLVD 601
1121 AVMTQVSRLASPLPPQASNAHDCDDSMGYQVFTLRVVKDGNCAWCPVRFRCGCK 1180
602 AVMTQVSRLASPLPPQASNAHDCDDSMGYQVFTLRVVKDGNCAWCPVRFRCGCK 661
1181 IDCEDRAFTIGNAYIAVDMPTALHLYQTSQBRVVDHESVEQSRAQAEPINLDSCLR 1240
662 IDCEDRAFTIGNAYIAVDMPTALHLYQTSQBRVVDHESVEQSRAQAEPINLDSCLR 721
1241 AFTSEELGENEMYCKCKTHCLAKLKLWLPILILHLKRFQVNGRWIKSQIVK 1300
722 AFTSEELGESEMYCKCKTHCLAKLKLWLPILILHLKRFQVNDWIKSQIVK 781

QY 1301 FPRESDPSAFLVPRDPALCOHKPLTPQDSELSERPRILAREVKKYDAQSSAGEEDVLLSK 1360
DB 782 FLRESFDPASFLVPRDPALCOHKPLTPQDSELSERPRILAREVKKYDAQSSAGKEDMLLSK 841
QY 1361 SPSSLSANISSPKGSPSSSRKSGTSCFSSKNSSPNSPRTLGRSKGLRLPQIGSKNKL 1420
DB 842 SPSSLSANISSPKGSPSSSRKSGTSCFSSKNSSPNSPRTLGRSKGLRLPQIGSKNKP 901
QY 1421 SSSKENLDASKENGAGQICELADALSRGHVILGGSQPELVTPQDHEVALANGFLYEHEACG 1480
DB 902 SSSKENLDASKENGAGQICELADALSRGHVILGGSQPELVTPQDHEVALANGFLYEHEACG 961
QY 1481 N----GYSNGOLGNHSEEDSTDDQEDTRIKPIYNYAIASHGILGGHVTYAKNPNC 1536
DB 962 NCGGNGYNGOLGNHSEEDSTDDQEDTRIKPIYNYAIASHGILGGHVTYAKNPNC 1021
QY 1537 KWCYNDSSCKELHPDEIDTDSAYILFYEQQIDYAOFLPKTIDGKKMADTSSMDEDFSD 1596
DB 1022 KWCYNDSSCKELHPDEIDTDSAYILFYEQQIDYAOFLPKTIDGKKMADTSSMDEDFSD 1081
QY 1597 YKYCVLQ 1604
DB 1082 YEKYSMLQ 1089
RESULT 9
ABG23332
ID ABG23332 standard; Protein; 1089 AA.
XX AC ABG23332;
XX DT 18-FEB-2002 (first entry)
XX XX Novel human diagnostic protein #23323.
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX XX WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX XX (HYSE-) HYSEQ INC.
XX XX Drmanac RT, Liu C, Tang YT;
XX XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS87519.
XX XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX PS Claim 20; SEQ ID No 53691; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX	Sequence	1089 AA;
XX	Query Match	52.9%; Score 4531; DB 22; Length 1089;
XX	Best Local Similarity	93.7%; Pred. No. 0;
XX	Matches	851; Conservative 22; Mismatches 31; Indels 4; Gaps 1;
QY	701	EVYKDMSPWPEMSPIANSKIDRHKVPTKGTGATGLNLGNTCFMNSSIQCVSNTPQLTQ 760
DB	182	EVYKDMSPWPEMSPIANSKIDRHKVPTKGTGATGLNLGNTCFMNSSIQCVSNTPQLTQ 241
QY	761	YFISGRHLYELNRTNPIGKMGHMAKCYGDLVQELWSGTOKNVAPLKLRWTIAKYAPFNG 820
DB	242	YFISGRHLYELNRTNPIGKMGHMAKCYGDLVQELWSGTOKNVAPLKLRWTIAKYAPFNG 301
QY	821	FQOQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSQRPDWEVAEAWDNLHNRNSIIV 880
DB	302	FQOQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSQRPDWEVAEAWDNLHNRNSIIV 361
QY	881	DLFHGQLRSQVKCTCGHISVREDFNLSPLPMDSVHLEITVIKLDGTTTPRYGLRL 940
DB	362	DLFHGQLRSQVKCTCGHISVREDFNLSPLPMDSVHLEITVIKLDGTTTPRYGLRL 421
QY	941	NWDEKYTKLQKLDLGLNSEQILLAEVHGSNIKFPQKNQKVLRSVSGFLCAFEIPVP 1000
DB	422	NWDEKYTKLQKLDLGLNSEQILLAEVHGSNIKFPQKNQKVLRSVSGFLCAFEIPVP 481
QY	1001	VSPISASPTQTDSSSPSTNEMTLTNGDLPRPIIPNGMPNTVVPCTGKNTFNGMV 1060
DB	482	VSPISASPTQTDSSSPSTNEMTLTNGDLPRPIIPNGMPNTVVPCTGKNTFNGMV 541
QY	1061	NGHMPSLPDSFTGYIIAVHKKWRTLYFLSSOKNRPFLFGMPLIVPCVTHRKDLYD 1120
DB	542	NGHMPSLPDSFTGYIIAVHKKWRTLYFLSSOKNRPFLFGMPLIVPCVTHRKDLYD 601
QY	1121	AVTQVSRLASPLPQBRASNAHQDCDSMGVQYPTLTVVQKNSCAWCPYRFRGCK 1180
DB	602	AVTQVSRLASPLPQBRASNAHQDCDSMGVQYPTLTVVQKNSCAWCPYRFRGCK 661
QY	1181	IDCEDRAFINAYIADWDPTALHAYQTSOERVDHESVOSRAOAEPLNLSCLR 1240
DB	662	IDCEDRAFINAYIADWDPTALHAYQTSOERVDHESVOSRAOAEPLNLSCLR 721
QY	1241	AFTSEELGENEMYCKCKTHCLATKCLDLRLFPILIIHLKRFQVNGRWIKSQIVK 1300
DB	722	AFTSEELGENEMYCKCKTHCLATKCLDLRLFPILIIHLKRFQVNGRWIKSQIVK 781
QY	1301	FPRSPDPSAFLVPRDPALCOHKPLTPQDELSPRILAREVKVDAQSGAEDVLLSK 1360
DB	782	FPRSPDPSAFLVPRDPALCOHKPLTPQDELSPRILAREVKVDAQSGAEDVLLSK 841
QY	1361	SPSLSANIISSPKGSPSSSKGTSQSPSSKNSSPNSPRTLGRSKGLRLPOIGSNKL 1420
DB	842	SPSLSANIISSPKGSPSSSKGTSQSPSSKNSSPNSPRTLGRSKGLRLPOIGSNKP 901
QY	1421	SSSKENLDASVENGAGQICELADALSRLHVGSGQPELVTPQDHEVALANGFLYEHAOG 1480
DB	902	SSSKENLDASVENGAGQICELADALSRLHVGSGQPELVTPQDHEVALANGFLYEHAOG 961
QY	1481	N----GYSNGQLGNHSDSTDDRETRIKPIYNLVAYISCHSGILGGHYVYAKNPNC 1536
DB	962	NGCGGVGNGQLGNHSDSTDDRETRIKPIYNLVAYISCHSGILGGHYVYAKNPNC 1021
QY	1537	KWYCYNDSSCKELHPEIDTDSAYILFYEQGIDYAOFLPKTDGKQWADTSSDDEDFESD 1596
DB	1022	KWYCYNDSSCKELHPEIDTDSAYILFYEQGIDYAOFLPKTDGKQWADTSSDDEDFESD 1081
QY	1597	YKTCVLTQ 1604
DB	1082	YKTCVLTQ 1089

RESULT 11

ABG21348
ID ABG21348 standard; Protein; 1125 AA.

XX AC ABG21348;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #21339.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PT 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR N-PSDB; AAS85535.

XX New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity

Claim 20; SEQ ID No 51707; 103pp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG00010-ABG03077 represent novel human
diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1125 AA;

Query Match 51.8%; Score 4433.5; DB 22; Length 1125;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 842; Conservative 15; Mismatches 49; Indels 7; Gaps 4;

QY 701 EVYKDMSPWPEMSPIANSKIDRHKVPTKGTGATGLNLGNTCFMNSSIQCVSNTPQLTQ 760

DB 220 EVYKDMSPWPEMSPIANSKIDRHKVPTKGTGATGLNLGNTCFMNSSIQCVSNTPQLTQ 279

QY 761 YFISGRHLYELNRTNPIGKMGHMAKCYGDLVQELWSGTOKNVAPLKLRWTIAKYAPFNG 820

280	YFISGRHLYELNRTNP	IGMKGHMAKCYGDLVQELWSGTQKVAPLKLRTIAKYAPKFDG	339	XX	WO200175067-A2.		
821	FOQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSGRDPDWEVAEAWDNHLRNRISIV	880	XX	11-OCT-2001.			
340	FOQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSGRDPDWEVAEAWDNHLRNRISIV	399	XX	30-MAR-2001; 2001WO-US08631.			
881	DLPHGQLRSQVKCTCGHISVRPDPNFSLPLPMDSYMHEITVIKLDGTTVPYRGLRL	940	XX	31-MAR-2000; 2000US-0540217.			
400	DLPHGQLRSQVKCTCGHISVRPDPNFSLPLPMDSYMHEITVIKLDGTTVPYRGLRL	459	XX	23-AUG-2000; 2000US-0649167.			
941	NMDEKTYGLKQLSDLCGLNSEQILLAEVHGSNIKNFPDQNKVRLSVSGFLCAFEI	1000	XX	(HYSE-) HYSEQ INC.			
460	NMDEKTYGLKQLSDLCGLNSEQILLAEVHGSNIKNFPDQNKVRLSVSGFLCAFEI	519	XX	Dmanac RT, Liu C, Tang YT;			
1001	VSPISASSPTQTDFFSSPSPTNEMFTLTITNGDLPRPIFIPNGMPTVVP	1060	XX	WPI; 2001-639362/73.			
520	VSPISASSPTQTDFFSSPSPTNEMFTLTITNGDLPRPIFIPNGMPTVVP	579	XX	N-PSDB; AAS85531.			
1061	NGHMPSLPDPTGYIIAHRKQMRTELYFLSQKRPRLFCMPLIVPCTVHTRKDL	1120	XX	New isolated polynucleotide and encoded polypeptides, useful in			
580	NGHMPSLPDPTGYIIAHRKQMRTELYFLSQKRPRLFCMPLIVPCTVHTRKDL	639	XX	diagnostics, forensics, gene mapping, identification of mutations			
1121	AVMIQVSRSLASPLPQESNAHQDCDSMGYQYPTFLRVVQXGNSCAWCPWYR	1180	XX	responsible for genetic disorders or other traits and to assess			
640	AVMIQVSRSLASPLPQESNAHQDCDSMGYQYPTFLRVVQXGNSCAWCPWYR	699	XX	biodiversity			
1181	IDGCEBRAFTGNAYIAVDWPTALHURYQTSQRRVDEHESVEQRRQAQAE	1240	XX	Claim 20; SEQ ID No 51703; 103pp; English.			
700	IDGCEBRAFTGNAYIAVDWPTALHURYQTSQRRVDEHESVEQRRQAQAE	759	XX	The invention relates to isolated polynucleotide (I) and			
1241	AFTSEBELGENEMYCSKCTHCLATKLDLWRLPPLIIHLKRFQVNGRWIKSQIVK	1300	XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,			
760	AFTSEBELGENEMYCSKCTHCLATKLDLWRLPPLIIHLKRFQVNGRWIKSQIVK	819	XX	polymerase chain reaction (PCR) primers, oligomers, and for chromosome			
1301	FPRESFDPALVPRDPALCOHKLTPQGDDELSEPRILAREVKKVDQAQSSAGE	1360	XX	and gene mapping, and in recombinant production of (II). The			
820	FPRESFDPALVPRDPALCOHKLTPQGDDELSEPRILAREVKKVDQAQSSAGE	879	XX	polynucleotides are also used in diagnostics as expressed sequence tags			
1361	SPSSLSANITSSPKGSPSSRSKSGTSCPSKNSPNSPRTL-GRSKGLRLPQIGSKNK	1419	XX	for identifying expressed genes. (I) is useful in gene therapy techniques			
880	SPSSLSANITSSPKGSPSSRSKSGTSCPSKNSPNSPRTL-GRSKGLRLPQIGSKNK	939	XX	to restore normal activity of (II) or to treat disease states involving			
1420	LSSKENLDASKENGAGQICELADALSRGHVLG-GSQPELVTPODHEVALANGFLYEHEA	1478	XX	CC (II). (II) is useful for generating antibodies against it, detecting or			
940	PSSSKKNDLASKENGAGQICELADALSRGHMRGVGSQPELVTPODHDVSLANGFLCEPEP	999	XX	quantitating a polypeptide in tissue, as molecular weight markers and as			
1479	CGN---GYSNGQLGNHSEEDSTDQREDTRIKPIYNYLVAISCHSILGGHYVYAKNP	1534	XX	a food supplement. (II) and its binding partners are useful in medical			
1000	CGNGCGGSGNGQLGNPSEARTDDQREDTHIXPIYNYLVAISCHSILGGHYVYAKNP	1059	XX	imaging of sites expressing (II). (I) and (II) are useful for treating			
1535	NCKWYCYNDSCKELHPDEIDTDSAYILFVEQIGIDYAOFLPKTDGKQWADTSSMEDFE	1594	XX	disorders involving aberrant protein expression or biological activity.			
1060	NCKWYCYNDSCKELHPDEIDTDSAYILFVEQIGIDYAOFLPKTDGKQWADTSSMEDFE	1119	XX	The polypeptide and polynucleotide sequences have applications in			
1595	SD-YKK 1599		XX	diagnostics, forensics, gene mapping, identification of mutations			
1120	SGLYKK 1125		XX	and to produce other types of data and products dependent on DNA and			
			XX	amino acid sequences. ABG00010-ABG30377 represent novel human			
			XX	diagnostic amino acid sequences of the invention.			
			XX	Note: The sequence data for this patent did not appear in the printed			
			XX	specification, but was obtained in electronic format directly from WIPO			
			XX	at ftp.wipo.int/pub/published_pct_sequences.			
			XX				
			XX	Sequence 1388 AA;			
			XX				
			XX	Query Match 50.5%; Score 4328.5; DB 22; Length 1388;			
			XX	Best Local Similarity 90.0%; Pred. No. 0;			
			XX	Matches 817; Conservative 21; Mismatches 29; Indels 41; Gaps 2;			
QY	701	EVRNKMDSWEEMSFANSKIDRHKVPTKGTGLSNLGNCTCFMNSIQCVSNQTLPTQ	760	QY	701	EVRNKMDSWEEMSFANSKIDRHKVPTKGTGLSNLGNCTCFMNSIQCVSNQTLPTQ	760
DB	518	EVRNKMDSWEEMSFANSKIDRHKVPTKGTGLSNLGNCTCFMNSIQCVSNQTLPTQ	577	DB	518	EVRNKMDSWEEMSFANSKIDRHKVPTKGTGLSNLGNCTCFMNSIQCVSNQTLPTQ	577
QY	761	YFISGRHLYELNRTNP	820	QY	761	YFISGRHLYELNRTNP	820
DB	578	YFISGRHLYELNRTNP	637	DB	578	YFISGRHLYELNRTNP	637
QY	821	FOQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSGRDPDWEVAEAWDNHLRNRISIV	880	QY	821	FOQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSGRDPDWEVAEAWDNHLRNRISIV	880
DB	638	FOQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSGRDPDWEVAEAWDNHLRNRISIV	697	DB	638	FOQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSGRDPDWEVAEAWDNHLRNRISIV	697
QY	881	DLPHGQLRSQVKCTCGHISVRPDPNFSLPLPMDSYMHEITVIKLDGTTVPYRGLRL	940	QY	881	DLPHGQLRSQVKCTCGHISVRPDPNFSLPLPMDSYMHEITVIKLDGTTVPYRGLRL	940
DB	698	DLPHGQLRSQVKCTCGHISVRPDPNFSLPLPMDSYMHEITVIKLDGTTVPYRGLRL	757	DB	698	DLPHGQLRSQVKCTCGHISVRPDPNFSLPLPMDSYMHEITVIKLDGTTVPYRGLRL	757
QY	941	NMDEKTYGLKQLSDLCGLNSEQILLAEVHGSNIKNFPDQNKVRLSVSGFLCAFEI	1000	QY	941	NMDEKTYGLKQLSDLCGLNSEQILLAEVHGSNIKNFPDQNKVRLSVSGFLCAFEI	1000
DB	758	NMDEKTYGLKQLSDLCGLNSEQILLAEVHGSNIKNFPDQNKVRLSVSGFLCAFEI	792	DB	758	NMDEKTYGLKQLSDLCGLNSEQILLAEVHGSNIKNFPDQNKVRLSVSGFLCAFEI	792
QY	1001	VSPISASSPTQTDFFSSPSPTNEMFTLTITNGDLPRPIFIPNGMPTVVP	1060	QY	1001	VSPISASSPTQTDFFSSPSPTNEMFTLTITNGDLPRPIFIPNGMPTVVP	1060

RESULT 12

ABG21344

ID ABG21344 standard; Protein; 1388 AA.

XX AC

XX ABG21344;

XX AC

XX 18-FEB-2002 (first entry)

XX DE

XX Novel human diagnostic protein #21335.

XX KW

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX XX

OS Homo sapiens.

Db 793 -----DSSSPSTNGMTLITNGDLPRPIIPNGMPNTVVPCTGKRNFTNGMV 840
Qy 1061 NGHMPSLPDSPTGYIIAVHKKMRTELYFLSSQKRPISLFGMPLIVPCTVHTRKKDLYD 1120
Db 841 NGHMPSLPDSPTGYIIAVHKKMRTELYFLSQENRPSLFGMPLIVECTVHTRKKDLYD 900
Qy 1121 AVMIQVSELASPLPOEASNAHQCDSDMGQYQFTLRVVKQDGNSCAWCPWTRFCRGCK 1180
Db 901 AVMIQVSWLAPLPPOEASIIHQARDNMGQYQFTLRVVKQDGNSCAWCPWTRFCRGCK 960
Qy 1181 IDCGEDRAFIAGNIAYVDWPTALHLRYQTSQERVDHESVEQSRRAQAEPIINLDSCLR 1240
Db 961 IDCGEDRAFIAGNIAYVDWPTALHLRYQTSQERVDHESVEQSRRAQAEPIINLDSCLR 1020
Qy 1241 AFTSEELGENEMYCCKKTHCLATKLDLWRLPPILIIHLKRFQVNGRWIKSQKIVK 1300
Db 1021 AFTSEELGESEMYCCKKTHCLATKLDLWRLPPILIIHLKRFQVNDQWIKSQKIVR 1080
Qy 1301 FPRSPFSAFLVPRDPALCOHKLTPQGBELSEPRILAREVKVDAQSSAGGEDVLLSK 1360
Db 1081 FLRESFSAFLVPRDPALCOHKLTPQGBELSEPRILAREVKVDAQSSAGGEDVLLSK 1140
Qy 1361 SPSSLSANIISPPKSPSSSRKSGTSCPSKNSPNSPRTLGSKGRLRLPOIGSKNKL 1420
Db 1141 SPSSLSANISSPPKSPSSSRKSGTSCPSKNSPNSPRTLGSKGRLRLPOIGSKNKP 1200
Qy 1421 SSSKENLDASKENGAGOCIELADALSRRHVLGGSOPELVTPQDHEVALANGFLYEHAAG 1480
Db 1201 SSSKKNLDASKENGAGOCIELADALSRRHVGSGSOPELVTPQDHEVALANGFLYEHAAG 1260
Qy 1481 N-----GYSNGOLGNSBEDSDTDQRETRIKPIYLNLYAISCBSGILGGHYVYAKNPNC 1536
Db 1261 NGCGDGYNGOLGNSBEDSDTDQRETHIKPIYLNLYAISCBSGILGGHYIYAKNPNC 1320
Qy 1537 KWCYNDSSCKELHPDEIDTDSAVILYEQOGLDYAQFLPKTDGKWADTSSWDEDFESD 1596
Db 1321 KWCYNDSSCKELHPDEIDTDSAVILYEQOGLDYAQFLPKIDGKWADTSSDEDSSED 1380
Qy 1597 YKCYCVLQ 1604
Db 1381 YEKYSMLQ 1388

RESULT 13
ID AAE19177
AC AAE19177 standard; Protein; 582 AA.
XX AAE19177;
XX AAE19177;
DT 21-MAY-2002 (first entry)
XX Human protease, PRTS-14 protein.

Human; protease; PRTS-14; enzyme; gastritis; cirrhosis; Crohn's disease; gastrointestinal disorder; autoimmune; inflammatory; cell proliferative; cardiovascular; developmental; epithelial; neurological; reproductive; AIDS; Acquired Immune Deficiency Syndrome; allergy; rheumatoid arthritis; anaemia; asthma; atherosclerosis; hypertension; myocardial infarction; hepatitis; cancer; psoriasis; Cushing's syndrome; hypothyroidism; eczema; epilepsy; Alzheimer's disease; Huntington's disease; Parkinson's disease; Pick's disease; infertility; vitiligo; drug screening; gene therapy.

OS Homo sapiens.
XX WO200208396-A2.
XX 31-JAN-2002.
XX 17-JUL-2001; 2001WO-US22397.
XX 21-JUL-2000; 2000US-220063P.
XX 28-JUL-2000; 2000US-221680P.

PR 04-AUG-2000; 2000US-223544P.
PR 11-AUG-2000; 2000US-224717P.
PR 16-AUG-2000; 2000US-225988P.
PR 23-AUG-2000; 2000US-227568P.
XX (INCYTE GENOMICS INC.
XX Deleageane AM, Gandhi AR, Hafalia AJA, Lu DAM, Patterson C;
PI Tribouley CM, Das D, Kallik DA, Nguyen DB, Lee EA, Khan FA;
PI Yue H, Au-Young J, Griffin JA, Policky JL, Ramkumar J, Yang J;
PI Thangavelu K, Ding L, Kearney L, Baughn MR, Borowsky ML;
PI Sanjanwala MS, Yao MG, Burford N, Walia NK, Lal P, Lee S, Todd S;
PI Lo TP, Tang YT, Elliott VS, Azimzai Y, Lu Y;
XX WPI; 2002-206082/26.
DR N-PSDB; AAD30581.
XX New human protease polypeptide, useful in diagnosis, prevention and
PT treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory,
PT cell proliferative, developmental, epithelial and neurological
PT disorders -
XX Claim 1; Page 152-153; 182pp; English.
PS The invention relates to an isolated human protease polypeptide (PRTS).
XX PRTS protein and DNA are useful for diagnosing, treating and preventing
CC gastrointestinal disorders (gastritis, cirrhosis, Crohn's disease),
CC autoimmune/inflammatory disorders (AIDS, allergy, rheumatoid arthritis,
CC anaemia, asthma), cardiovascular disorder (atherosclerosis, hypertension,
CC myocardial infarction), cell proliferative disorders (hepatitis, cancer,
CC psoriasis), developmental disorders (Cushing's syndrome, hypothyroidism),
CC epithelial disorder (vitiligo, keloid, eczema), neurological disorders
CC (epilepsy, Alzheimer's disease, Pick's disease, Huntington's disease,
CC Parkinson's disease), and reproductive disorders (infertility). PRTS
CC protein is useful in a number of drug screening techniques and to
CC analyse the proteome of a tissue or cell type. PRTS DNA is useful for
CC creating knockin humanised animals or transgenic animals to model human
CC diseases, in somatic or germline gene therapy and in microarrays
CC utilising fluids or tissues from patients to detect altered PKIN
CC expression. The present sequence is human PRTS-14 protein.
XX Sequence 582 AA;

Query Match 36.6%; Score 3138; DB 23; Length 582;
Best Local Similarity 100.0%; Pred. No. 2.6e-270;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1023 MFTLTNGDLPRPIIPNGMPNTVVPCTGKRNFTNGMVHMPSLPDSPTGYIIAVHKK 1082
Db 1 MFTLTNGDLPRPIIPNGMPNTVVPCTGKRNFTNGMVHMPSLPDSPTGYIIAVHKK 60

QY 1083 MWRTELYFLSSQKRPISLFGMPLIVPCTVHTRKKDLYDVAWMIQVSRLASPLPPOEASNAH 1142
Db 61 MWRTELYFLSSQKRPISLFGMPLIVPCTVHTRKKDLYDVAWMIQVSRLASPLPPOEASNAH 120
QY 1143 QCDSDMGQYQFTLRVVKQDGNSCAWCPWTRFCRGCKIDCGEDRAFIAGNIAYVDWPT 1202
Db 121 QCDSDMGQYQFTLRVVKQDGNSCAWCPWTRFCRGCKIDCGEDRAFIAGNIAYVDWPT 180
QY 1203 ALHLRYQTSQERVDHESVEQSRRAQAEPIINLDSCLRFTSEELGENEMYCCKKTH 1262
Db 181 ALHLRYQTSQERVDHESVEQSRRAQAEPIINLDSCLRFTSEELGENEMYCCKKTH 240
QY 1263 CLATKKLDLWRLPPILIIHLKRFQVNGRWIKSQKIVKFPRESFDPFAFLVPRDPALCOH 1322
Db 241 CLATKKLDLWRLPPILIIHLKRFQVNGRWIKSQKIVKFPRESFDPFAFLVPRDPALCOH 300
QY 1323 KPLTPQGBELSEPRILAREVKVDAQSSAGGEDVLLSKSPSSLSANIISPPKSPSSSRK 1382
Db 301 KPLTPQGBELSEPRILAREVKVDAQSSAGGEDVLLSKSPSSLSANIISPPKSPSSSRK 360
QY 1383 SGTSCPSKNSPNSPRTLGSKGRLRLPOIGSKNKLSSKENLDASKENGAGOCIELA 1442

Db	361	SGTSCPSKSNPSNPSRTTGRSKGRRLRPQIGSKVKLSSSKENIDASKENGAGQICELA	1502
QY	1443	DALSRGHVLGSGQPELVTPODHEVALANGFLYEHEACGNGYNGQLGNESEDSTDDQRE	1502
Db	421	DALSRGHVLGSGQPELVTPODHEVALANGFLYEHEACGNGYNGQLGNESEDSTDDQRE	480
QY	1503	DRRIKPIYNLVAISCHSGILGGGHYTYAKNPCKWYCYNDSSCKELHPDIDTDSAYIL	1562
Db	481	DRRIKPIYNLVAISCHSGILGGGHYTYAKNPCKWYCYNDSSCKELHPDIDTDSAYIL	540
QY	1563	FYEQQGIDYAQFLPKTGTGKMAWDTSSMDEDFESDYKKYCVLQ	1604
Db	541	FYEQQGIDYAQFLPKTGTGKMAWDTSSMDEDFESDYKKYCVLQ	582
RESULT 14			
XX	ABB71649	standard; Protein; 1779 AA.	
XX	ABB71649;		
XX	DT	26-MAR-2002 (first entry)	
XX	DE	Drosophila melanogaster polypeptide SEQ ID NO 41739.	
XX	KW	Drosophila; developmental biology; cell signalling; insecticide;	
XX	KW	pharmaceutical.	
XX	OS	Drosophila melanogaster.	
XX	XX	WO200171042-A2.	
XX	PD	27-SEP-2001.	
XX	PF	23-MAR-2001; 2001WO-US09231.	
XX	XX	23-MAR-2000; 2000US-191637P.	
PR	PR	11-JUL-2000; 2000US-0614150.	
XX	PA	(PEKE) PE CORP NY.	
PI	PI	Venter JC, Adams M, Li PWD, Myers EW;	
XX	XX	WPI; 2001-656860/75.	
DR	DR	N-PSDB; ABL15752.	
XX	PT	New isolated nucleic acid detection reagent for detecting 1000 or more	
PT	PT	genes from Drosophila and for elucidating cell signalling and cell-cell	
PT	PT	interactions -	
XX	PS	Disclosure; SEQ ID NO 41739; 2lpp + Sequence Listing; English.	
XX	XX	The invention relates to an isolated nucleic acid detection reagent	
CC	CC	capable of detecting 1000 or more genes from Drosophila. The invention is	
CC	CC	useful in developmental biology and in elucidating cell signalling and	
CC	CC	cell-cell interactions in higher eukaryotes for the development of	
CC	CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
CC	CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA	
CC	CC	sequences (ABL01840-ABL16175) and the encoded proteins	
CC	CC	(ABB57737-ABB72072).	
CC	CC	The sequence data for this patent did not form part of the printed	
CC	CC	specification, but was obtained in electronic format directly from WIPO	
CC	CC	at ftp.wipo.int/pub/published_pct_sequences.	
XX	XX	Sequence 1779 AA;	
Query Match 35.3%; Score 3019.5; DB 22; Length 1779;			
Best Local Similarity 35.8%; Pred. No. 8e-259;			
Matches 685; Conservative 283; Mismatches 475; Indels 471; Gaps 46;			
2Y	21	TDVELRLKDAFKETCLGS-YYNGQCHPTREVLGCVPPKVAEVLYCSFGTSGKLHFN 79	
Db	1	SDVELRLRDAFKSAGVGNFVSRNAFOQDLVCEGVPPKIVDMLYACGGTQRGISFND 60	

QY 1002 -----SPISASSP----- 1009
Db 1091 DEAGKDELAPORNCYRMIYREWFCECEIRQESASVCFSIESLILSSSPENTFMH 1150
QY 1010 -----TOTDRSSSTNEMPTLTITNGDLPRIP-----NGMPT 1045
Db 1151 GAAAOQKRVSSAKLHTESTNISMGTTHSGENSWESSLTPEIPLADLEVPVSSRNGSGGE 1210
QY 1046 VPCGTEKFTNGMVGHM--PSL-----PDSPTG----- 1074
Db 1211 DCSYRTSPNDSSGLSTGHTLGASLDVDEQAEGBNAEDHDQDQITTSQPETSSGVSSRS 1270
QY 1075 -----YIIVAKRMVRELVELSSQKRPISLFGMLPLVPCVTRKDLDAVWQ 1125
Db 1271 SOPPKAKILVAVARKITRDSYFELS YHKTRPSLFGVPLIPNSEGTHKDLCAVWLQ 1330
QY 1126 VSLASPLP-POEASNHQDCCDDSGYQYPTFLRVQKDGNSCAWCPWYRCRCCKIDOG 1184
Db 1331 VSRLLSPLPATTEQANHAADCCDLSGLYDFPTLRAVKADGLTCAICPWSFCRCCEIRCN 1390
QY 1185 EDPAFIG-----NA-----YI 1195
Db 1391 NDYVLOGALPINAASNTSTPKNAKPPSLNLEAKTPEYTAASLSVTPTTKYFEDFTI 1450
QY 1196 AVMDPTALHLYQTSQBRVDEHESVQSRRAQAEPINLDSCLRAFTSBEELGENEMY 1255
Db 1451 AIDMDPTALHLYQSTLERLWVDHETIAISRRQVEPVDLNHCLRAFTSEKL--EOWYH 1508
QY 1256 CSKCTHCLATKLDLWLPILIIHLKRFVNGRWIKSQIYKFPRESFPDPSAPL--V 1313
Db 1509 CSHCKGKPKATKGLQIMKLPILIVHLKRFVNGRWIKSQIYKFPRESFPDPSAPL--V 1568
QY 1314 PRDPALCOHKLPTQGDSESPRILAREVKVDAQSSAGEEDVLLSKSPSLSNISSP 1373
Db 1569 PQETIL-RHKEL-----EL-----XNDAMTWTATNEVV----- 1596
QY 1374 KSPSSSKSTSCPS--KNSSPNSPRTLRSGKRLRFPQIGSKNKLSSKENLDASK 1431
Db 1597 -----SELDEIDAPSKEVEELNQT-----GSKATASPPPTGNILRQSKTK----- 1639
QY 1432 ENGAGIICELADALSRGHVLGSGQPELVTPQDHEVALANGELYEACGNGYNGQGNH 1491
Db 1640 -----NAVROQLISTSTK--TP-----IVDGEFEDYH-----QH 1668
QY 1492 SEEDSTDQREDTRIKPTIYNAISCHSGILGGHYVYAKNPCKWYCVNDSSCKEL-H 1550
Db 1669 RLKPDVD-----QFDPYRLYAVVSHSGMLNGHYISYASNATGSGWYCVNDSSCREISQ 1722
QY 1551 PDEIDTDSAYILFVEQOQIDYAPLPTDGHKMDTSSMD---EDFESDYKCYC 1601
Db 1723 KPVIDPSAAYLLFYERKGLDYEPYLPNIEGRTLPNTASVPLEVDETEGELKLC 1776

RESULT 15
RAU87244
ID AAU87244 standard; Protein; 539 AA.
AC AAU87244;
XX
DT 05-JUN-2002 (first entry)
DE Novel central nervous system protein #154.
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasia; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischemia; angiodysplasia;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.

XX Homo sapiens.
XX WO200155318-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01332.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 14-JUL-2000; 2000US-0217496.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
XX 25-SEP-2000; 2000US-0234997.
XX 25-SEP-2000; 2000US-0234998.

26-SEP-2000; 2000US-0235484.
 27-SEP-2000; 2000US-0235834.
 27-SEP-2000; 2000US-0235836.
 29-SEP-2000; 2000US-0236327.
 29-SEP-2000; 2000US-0236367.
 29-SEP-2000; 2000US-0236368.
 29-SEP-2000; 2000US-0236369.
 29-SEP-2000; 2000US-0236370.
 02-OCT-2000; 2000US-0236802.
 02-OCT-2000; 2000US-0237037.
 02-OCT-2000; 2000US-0237038.
 02-OCT-2000; 2000US-0237039.
 02-OCT-2000; 2000US-0237040.
 13-OCT-2000; 2000US-0239335.
 13-OCT-2000; 2000US-0239337.
 20-OCT-2000; 2000US-0240960.
 20-OCT-2000; 2000US-0241221.
 20-OCT-2000; 2000US-0241785.
 20-OCT-2000; 2000US-0241786.
 20-OCT-2000; 2000US-0241787.
 20-OCT-2000; 2000US-0241808.
 20-OCT-2000; 2000US-0241809.
 20-OCT-2000; 2000US-0241826.
 01-NOV-2000; 2000US-0244617.
 08-NOV-2000; 2000US-0246474.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246478.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249219.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249264.
 17-NOV-2000; 2000US-0249265.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 17-NOV-2000; 2000US-0249300.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251388.
 05-DEC-2000; 2000US-0256719.
 08-DEC-2000; 2000US-0256719.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251858.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 03-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-581633/65.
 XX N-PSDB; ABK43574.
 DR New isolated nucleic acid encoding a protein for diagnosing,
 PT preventing, treating or ameliorating medical conditions and used as
 XX food additives or preservatives -
 PS Claim 9; SEQ ID No 762; 837pp; English.
 XX The invention describes an isolated nucleic acid molecule (I) encoding a
 CC novel central nervous system protein. (I) and polypeptides (III) encoded
 CC by (I), are used to treat a medical condition and in diagnosis of a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiodysplasia, nervous system disorders e.g. Alzheimer's disease and
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
 CC adenocarcinomas and irritable bowel syndrome, reproductive system
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
 CC acute kidney failure and blood related disorders e.g. myocardial
 CC infarction. The polypeptides can also be used to aid wound healing and
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
 CC maintain organs before transplantation, for supporting cell culture of
 CC primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 Query Match 31.7%; Score 2711; DB 22; Length 539;
 Best Local Similarity 93.3%; Pred. No. 3e-232;
 Matches 512; Conservative 9; Mismatches 6; Indels 22; Gaps 2;
 QY 117 ERLHVVDTGKVPDTLRKCFSEGEKVKYKFNFLNFKDAFTSRWLLSGGVVTLTDDS 176
 Db 7 ERLHVVDTGKVPDTLRKCFSEGEKVKYKFNFLNFKDAFTSRWLLSGGVVTLTDDS 66
 QY 177 DTPTFYOTLAGVTHLESDDIIEKRYWLLKAQSRGTRGFLETFGLVSPPIRPSISEGL 236
 Db 67 DTPTFYOTLAGVTHLESDDIIEKRYWLLKAQSRGTRGFLETFGLVSPPIRPSISEGL 126
 QY 237 FNAFDENRNDHIDFKETISCGLSACCGPLAERQKFCFKFVDFVDRDGVLSRVELRDMVVAL 296
 Db 127 FNAFDENRNDHIDFKETISCGLSACCGPLAERQKFCFKFVDFVDRDGVLSRVELRDMVVAL 186
 QY 297 LEVWKDNRDIDIELEHMDLSDIIEGILNAHDTTGMHGLTLEDYQIWSKVLANEFNL 356
 Db 187 LEVWKDNRDIDIELEHMDLSDIIEGILNAHDTTGMHGLTLEDYQIWSKVLANEFNL 246
 QY 357 FQVCHIVILGLRPATPEEGQIIRGWLERSRYGLQAGHNWFIISMQWQKQKYYKYDAN 416
 Db 247 FQVCHIVILGLRPATPEEGQIIRGWLERSRYGLQAGHNWFIISMQWQKQKYYKYDAN 306
 QY 417 FVWIEPSSVILNGKYSFGTAHMEQVEDRIGSSLSYVNTTEKFSNISTASESETAG 476
 Db 307 FVWIEPSSVILNGKYSFGTAHMEQVEDRIGSSLSYVNTTEKFSNISTASESETAG 366
 QY 477 SGFLYSATPGADVCFARQHTNSDNNOCILGANGNILLNLPKPGADNQLVTPQEPVK 536
 Db 367 SGFLYSATPGADVCFARQHTNSDNNOCILGANGNILLNLPKPGADNQLVTPQEPVK 426
 QY 537 ATSLTLEGGRLKRTPTQLIHGRDYEMVPEPVWRALYHWYGANLALPRPVIKNSKTDIPELE 596
 Db 427 ATSLTLEGGRLKRTPTQLIHGRDYEMVPEPVWRALYHWYGANLALPRPVIKNSKTDIPELE 486

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:08:55 ; Search time 22 seconds
(without alignments)
3044.846 Million cell updates/sec

Title: US-09-888-615-73

Perfect score: 8563

Sequence: 1 NGAKESRICFLSYEALRRV.....DTSSMDPESDYKVCVLQ 1604

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310859 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/aaa/5A COMB.pep:*
2: /cgn2_6/prodata/1/aaa/5B COMB.pep:*
3: /cgn2_6/prodata/1/aaa/6A COMB.pep:*
4: /cgn2_6/prodata/1/aaa/6B COMB.pep:*
5: /cgn2_6/prodata/1/aaa/PCITUS COMB.pep:*
6: /cgn2_6/prodata/1/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2392.5	27.9	462	4	US-09-166-350-18
2	529	6.2	292	4	US-09-406-045-5
3	367	4.3	926	1	US-08-159-340A-2
4	347.5	4.1	193	1	US-08-159-340A-3
5	345	4.0	63	1	US-08-159-340A-5
6	334	3.9	67	1	US-08-159-340A-6
7	311	3.6	55	3	US-09-019-095A-32
8	286	3.3	54	1	US-08-159-340A-8
9	280	3.3	52	1	US-08-159-340A-7
10	217.5	2.5	1123	4	US-09-408-865-1
11	215	2.5	545	3	US-09-019-095A-38
12	201.5	2.4	509	3	US-09-019-095A-9
13	200	2.3	508	3	US-09-019-095A-8
14	200	2.3	521	3	US-09-019-095A-22
15	200	2.3	526	3	US-09-019-095A-2
16	184	2.1	418	3	US-09-019-095A-10
17	173	2.0	494	3	US-09-019-095A-24
18	160.5	1.9	220	4	US-09-399-913-24
19	160.5	1.9	220	4	US-09-298-731-24
20	156.5	1.8	191	3	US-08-655-352-5
21	156.5	1.8	191	3	US-08-655-352-6
22	156.5	1.8	191	4	US-09-258-016-5
23	156.5	1.8	191	4	US-09-258-016-6
24	156.5	1.8	191	4	US-09-257-825B-5
25	156.5	1.8	191	4	US-09-257-825B-6
26	156.5	1.8	220	4	US-09-399-913-26
27	156.5	1.8	220	4	US-09-298-731-26

28 155 1.8 291 4 US-09-167-206-22 Sequence 22, Appl
29 152.5 1.8 256 4 US-09-399-913-32 Sequence 32, Appl
30 152 1.8 118 3 US-08-905-223-399 Sequence 399, Appl
31 151.5 1.8 225 4 US-09-399-913-30 Sequence 30, Appl
32 151.5 1.8 225 4 US-09-298-731-30 Sequence 30, Appl
33 151.5 1.8 252 4 US-09-399-913-22 Sequence 22, Appl
34 151.5 1.8 252 4 US-09-399-913-28 Sequence 28, Appl
35 151.5 1.8 252 4 US-09-399-913-42 Sequence 42, Appl
36 151.5 1.8 252 4 US-09-298-731-22 Sequence 22, Appl
37 151.5 1.8 252 4 US-09-298-731-28 Sequence 28, Appl
38 151.5 1.8 257 4 US-09-399-913-16 Sequence 16, Appl
39 151.5 1.8 257 4 US-09-298-731-16 Sequence 16, Appl
40 151.5 1.8 270 4 US-09-399-913-18 Sequence 18, Appl
41 151.5 1.8 270 4 US-09-298-731-18 Sequence 18, Appl
42 151 1.8 193 3 US-08-655-352-4 Sequence 4, Appl
43 151 1.8 193 4 US-09-258-016-4 Sequence 4, Appl
44 151 1.8 193 4 US-09-257-825B-4 Sequence 4, Appl
45 150 1.8 174 1 US-08-328-322-17 Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-166-350-18
; Sequence 18, Application US/09166350A
; Patent No. 6440663
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Chen, Yao
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alex
; TITLE OF INVENTION: Renal Cancer Associated Antigens and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: L0461/7051
; CURRENT APPLICATION NUMBER: US/09/166,350A
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: US 09/166,350
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-166-350-18

Query Match 27.9%; Score 2392.5; DB 4; Length 462;
Best Local Similarity 97.4%; Pred. No. 9.2e-201;
Matches 453; Conservative 1; Mismatches 6; Indels 5; Gaps 2;

QY 536 KATSLTLEGGELKRTQLHGRDYEMVPPVWRALYHWTGANLALPRVVKSKTDIPEL 595
DB 1 KATSLTLEGGELKRTQLHGRDYEMVPPVWRALYHWTGANLALPRVVKSKTDIPEL 60
QY 596 ELFPYLLFLRQOPATRTQOSNIWVNMGNVPSPNAPLKRVLAYTGCFSRMQTIKEIHEYL 655
DB 61 ELFPYLLFLRQOPATRTQOSNIWVNMGNVPSPNAPLKRVLAYTGCFSRMQTIKEIHEYL 120
QY 656 SQRLRKEEDMRUWLNSENLYLTDDHDKLEYLKIODEQHLVIEVRNKMSPPEMSF 715
DB 121 SQRLRKEEDMRUWLNSENLYLTDDHDKLEYLKIODEQHLVIEVRNKMSPPEMSF 180
QY 716 IANSSKIDRHKVPTEKGATGLSNLGNCTFWNSSIQCVSNTQPLTQYFISGRHLYELNRN 775
DB 181 IANSSKIDRHKVPTEKGATGLSNLGNCTFWNSSIQCVSNTQPLTQYFISGRHLYELNRN 240
QY 776 PIGMKHMAKCYGDLVQELWSGTQKNVAPLKLRTWTAKYAPRNFNGQQDSQELLAFILD 835
DB 241 PIGMKHMAKCYGDLVQELWSGTQKNVAPLKLRTWTAKYAPRNFNGQQDSQELLAFILD 300

QY 836 GLHEDLNVRHVKPYVELKDSGRDPDWEVAABANDNHLRRNRSIIVDLFHQOLRSQVCKCT 895
DB 301 GLHEDLNVRHVKPYVELKDSGRDPDWEVAABANDNHLRRNRSIIVDLFHQOLRSQVCKCT 360
QY 896 CCHISVRDPDNFSLPLPMDSYMHLEITVTKLDTTPVRYGLRLNMDKTYGLKKQLSD 955
DB 361 CCHISVRDPDNFSLPLPMDSYMHLEITVTKLDTTPVRYGLRLNMDKTYGLKKQLSD 420
QY 956 LCGLNSEQILLAEVHGSNIKFPQDNQKVR--LSVSGFLCAFEIP 998
DB 421 LCGLNSEQILLAEVHGSNIKFPQDNQKVR--LSVSGFLCAFEIP 998
RESULT 2
US-09-406-045-5
; Sequence 5, Application US/09406045
; Patent No. 6451994
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23413, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-50
; CURRENT APPLICATION NUMBER: US/09/406,045
; CURRENT FILING DATE: 1999-09-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ProDom consensus sequence
US-09-406-045-5

Query Match 6.2%; Score 529; DB 4; Length 292;
Best Local Similarity 23.1%; Pred. No. 1.1e-37;
Matches 137; Conservative 52; Mismatches 80; Indels 324; Gaps 10;
QY 732 GATGLNLGNTCFMSSQCVSNTQPLQYFISGRHLYEL--NRTNPIMGMHAKYVGD 789
DB 9 GYTGLKLNLTGNTCMNSVLQCLYHVPPLREYFLEDEYSEVMNNSNPLGMKELATAYAK 68
QY 790 LVQELWS--GTOKNVAPLKLRTIAYAPFN--GFOQDSQELLAFLLDGLHEDLNVRHEK 847
DB 69 LVHQMNSNSKSVAPQLTTLTVGVKSPFSGYQQQSQEFLKQLDADHEDFNSLMEK 128
QY 848 PYVE--LKDSGKPD--MEVAEAENDNHLRRNRSIIVDLFHQOLRSQVCKCTCGHIS--VR 902
DB 129 PYVEEQVQDSNEKSTALNVVSEAEWENHKRNDISIITDIFQCFKSTIKCPSCHESTETT 188
QY 903 FQPFNLSPLPMDSYMHLEITVTKLDTTPVRYGLRLNMDKTYGLKKQLSDLCGLNSE 962
DB 189 FQPFNLSPLPMDSYMHLEITVTKLDTTPVRYGLRLNMDKTYGLKKQLSDLCGLNSE 962
QY 963 QILLAEVHGSNIKFPQDNQKVRLSVSGFLCAFEIPFVPSIASPTQTDFFSSSPSTNE 1022
DB 204 ----- 203
QY 1023 MFTLTNGDLPRPIFIPNGMNTVVPCTGTEKNTNGMNGMPSLPDPSFTGYIIAIVHRK 1082
DB 204 ----- 203
QY 1083 MMRTELYFLSSQKRPFLGFMPLIVPCTVHTRKDLVDVAVIQRSLASPLPQOASNHA 1142
DB 204 -----ADNHQ 208
QY 1143 QDCDDSMGYQYPTLVRVQKQNSCAWCPWYFCRGCKIDCGEDRAFIAGYAVIWDVDP 1202
DB 209 ----- 208
QY 1203 ALHLRYCTQERVVDEHSEVQSRRAQBPINLDSCLRAFTSEELGENEMYVCKSKKTH 1262
DB 209 -----NLQCLSEFTKTELEGDNKWCYCFKCKK 237

QY 1263 CLATKKLDLWRLPPLIILHLKRFQFVNGRWIKSOKI---VKFPRESFDPASFL 1312
DB 238 QEATKKLDLWRLPPLIILHLKRFQFVNGRWIKSOKI---VKFPRESFDPASFL 289
RESULT 3
US-08-159-340A-2
; Sequence 2, Application US/08159340A
; Patent No. 5565352
; GENERAL INFORMATION:
; APPLICANT: Hochstrasser, Mark
; APPLICANT: Papa, Feroz
; TITLE OF INVENTION: DEUBIQUITINATING ENZYME: COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,340A
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD:112/HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 926 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-159-340A-2

Query Match 4.3%; Score 367; DB 1; Length 926;
Best Local Similarity 21.8%; Pred. No. 1.5e-22;
Matches 159; Conservative 100; Mismatches 225; Indels 242; Gaps 30;
QY 242 ENRDNHDIPKEISCGLSACRCGRPLAERQKFCVDPVDRDGVLSRVLELDVMVALLEVWK 301
DB 219 ENKSHDITKHIIC-----LEPISFKMSYSDHD-----LEKSLI----- 253
QY 302 DNRDTPDIPELHMDLS--DIVEGILNAHDTTKMGHLETDYQIWSVKNVLANEFNLFLQVC 360
DB 254 ---TSNSEIKMFQSRNLFKFIILYDA-----NEYNV-KQOSVLLDILVNSHFE-- 299
QY 361 HVLGLRPAATPE-----BEGQITGWLERSRYGLQAGHNVFIISMQWQWQKEYVKY 413
DB 300 -----KPISDDTKFIILESG--PPGWL--KSNYGRQVSSS-----F 332
QY 414 DAMPVVIEPSSVLNGKYSFGTAHAPMEQVEODRIGSS-----LSYVNTTEE 459
DB 333 PSNNIKDDSVIYNGN--TSGLSLQHLPKMSIRHSMDSDMKMLVAPTPLNHLQOQQQ 390
QY 460 KFDENISTASEAETAGSGFLYSATPGADVCFARQNTSDNNQCLLGANGNILLHLNPO 519
DB 391 QOSDNDHVLKRSS-----SFKKLFENYTPSPNPKNSNLYSISLSSISSPS 437
QY 520 KPGALDNPQVLTQEPVKATSLTLEGGRKLTQPLIHGRDVEYVPEVPVRALYHWYGANLA 579

Db 438 P-----LPLHSPDPVKGNSFRI---NYBPETHL-----W-----463
QY 580 LRPVINKNSTDIPPELEPRYLLELRQOPATEQOSNIWVWNGVPSNAPLKRVLAYT 639
Db 464 -----KNSBTD-----FVINOQELNHSFAHIAPIFKATSP-----497
QY 640 GCFSRMOT--IKEIHYLSORLAKEEDMRMLWLYNSENYLTLLDDHDKLEYLKIQDEQH 697
Db 498 ---SRTATPKLQRPPTQISMNLN-----MNSNGHSSATSTIQ-----531
QY 698 LVLEVNKOMSWEEMSFANSKSIDRHVK-PPEKG-----ATGLSNLGNTCFNMSSIQ 750
Db 532 -----PSCUS-LSNNDLSLHTDVTPTSSHNYDLDFAVGLNGLNSCYMNCIIQ 578
QY 751 CVSNTQPLQYFTISGRHLYELNRTNPIGMKGMAKCYGDLVQELW-----SGTK-NVAPL 805
Db 579 CILGTHLTIQIFLDDSYAKHININSLKSGILAKYFAALVHMVYKEQVDGSKKISISPI 638
QY 806 KLRWTIAYAPRNGFQQODSOELLAFLLDGLHEDLNVRVHEKPYVELKDSGRPDWEVAA 865
Db 639 KFKLACGSVNSLFTASQODCOFCOFLDGLHEDLNQCGSNP-----PLKELSQ 688
QY 866 EA-----WDNHLRRNRSIVVDLFGHQLRSQVCKCTCGHISVREDFPNFL 909
Db 689 EABARREKLSLRASSIEWERFLTDFSVIVDLFQCYASRLKRCVCSHTSTTYQFTVL 748
QY 910 SLPLP 914
Db 749 SIPI 753
RESULT 4
US-08-159-340A-3
; Sequence 3, Application US/08159340A
; Patent No. 5565352
; GENERAL INFORMATION:
; APPLICANT: Hochstrasser, Mark
; TITLE OF INVENTION: DEUBIQUITINATING ENZYME: COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,340A
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD:112/HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 193 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-159-340A-3

Query Match 4.1%; Score 347.5; DB 1; Length 193;
Best Local Similarity 36.5%; Pred. No. 4.2e-22;
Matches 74; Conservative 31; Mismatches 67; Indels 31; Gaps 4;
QY 733 ATGLSNLGNTCFNMSSIQCVSNTQPLQYFTISGRHLYELNRTNPIGMKGMAKCYGDLVQ 792
Db 1 AVGLNGLNSCYMNCIIQCIILGTHLTIQIFLDDSYAKHININSLKSGILAKYFAALVH 60
QY 793 ELW-----SGTK-NVAPLKLRTIAYAPRNGFQQODSOELLAFLLDGLHEDLNVRVHEK 847
Db 61 MMYKEQVDGSKKISISPIKFKLACGSVNSLFTASQODCOFCOFLDGLHEDLNQCGSN 120
QY 848 PYVELKDSGRPDWEVAAEA-----WDNHLRRNRSIVVDLFGHQLRSQV 891
Db 121 P-----PLKELSQEABARREKLSLRASSIEWERFLTDFSVIVDLFQCYASRL 170
QY 892 KCKTCGHISVREDFPNFLSLPLP 914
Db 171 KRCVCSHTSTTYQFTVL SIPI 193
RESULT 5
US-08-159-340A-5
; Sequence 5, Application US/08159340A
; Patent No. 5565352
; GENERAL INFORMATION:
; APPLICANT: Hochstrasser, Mark
; TITLE OF INVENTION: DEUBIQUITINATING ENZYME: COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,340A
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD:112/HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-159-340A-5
Query Match 4.0%; Score 345; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 8.7e-23;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 733 ATGLSNLGNTCFNMSSIQCVSNTQPLQYFTISGRHLYELNRTNPIGMKGMAKCYGDLVQ 792
Db 1 ATGLSNLGNTCFNMSSIQCVSNTQPLQYFTISGRHLYELNRTNPIGMKGMAKCYGDLVQ 60
QY 793 ELW 795
|||

;; CURRENT APPLICATION NUMBER: US/09/019,095A
;; CURRENT FILING DATE: 1998-02-05
;; PRIOR APPLICATION NUMBER: PCT/US96/12884
;; PRIOR FILING DATE: 1996-08-07
;; PRIOR APPLICATION NUMBER: US 60/002,066
;; PRIOR FILING DATE: 1995-08-09
;; PRIOR APPLICATION NUMBER: US 60/019,787
;; PRIOR FILING DATE: 1996-06-14
;; NUMBER OF SEQ ID NOS: 51
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 38
;; LENGTH: 545
;; TYPE: PRT
;; ORGANISM: murine
US-09-019-095A-38

Query Match 2.5%; Score 215; DB 3; Length 545;
Best Local Similarity 28.5%; Pred. No. 1.2e-09;
Matches 65; Conservative 32; Mismatches 79; Indels 52; Gaps 8;
QY 693 QDECHLVIEVRK---MSWPEMSFIANSKIDRHKVPTEKGATGLSLNLTGTCFNMSSI 749
Db 24 QDEAQQVVELTANDEKSLW-----ECQGPQC-GLQNTGNSCYLNAAL 66
QY 750 QCVSNTQPLTQYFISGRHLYELNRTNPIGMKGMKCYGDLVQELWSGTQKNVAPLKL 809
Db 67 QCLTHPTPLADYMLSOE--YSQCCSPGCK--MCAMEALVTQSLHSHSGDVMKPSQIL 122
QY 810 TIAYAPRFNGFQQDSQELAFLLDGLHEDLNVRVHEKPYVELKSDGPRDVEVAEAWD 869
Db 123 TSA-----FKHQQEDAHEFLMFTLETWHESCLQVHRSQETSDES----- 163
QY 870 NHLNRNRSIVVDLPHGQLRSQVKCKTCGHSVRFPDPFNFLSLPLPMS 917
Db 164 -----SPIHDFGGLWRSQIKCLCOGTSQSDTYD--RFLDPLDISS 202

RESULT 12

US-09-019-095A-9
; Sequence 9, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; TITLE OF INVENTION: Cell Growth
; FILE REFERENCE: DFCI-435p2A2
; CURRENT APPLICATION NUMBER: US/09/019,095A
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Murine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(509)
; NAME/KEY: VARIANT
; LOCATION: (1)...(509)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-019-095A-9

Query Match 2.4%; Score 201.5; DB 3; Length 509;
Best Local Similarity 31.6%; Pred. No. 1.6e-08;
Matches 60; Conservative 30; Mismatches 53; Indels 47; Gaps 10;

QY 735 GLSLNLTGTCFNMSSIQCVSNTQPLTQYFISGRHLYELNRT--NPIGMK-GHMAKCYGDLV 791
Db 41 GLQNTGNSCYLNAALQCLTHPTPLADYMLSOE-----SQCCSPGCKMCAMEEC---VT 93
QY 792 QELWSGTQKNVAPLKLRTIAK---YAPRFNGFQQDSQELAFLLDGLHEDLNVRVHEK 847
Db 94 QSL-----XLSLXGDVWKPSQILTSFAFHKQEDAHEFLMFTLETWHESCLQVH-- 142
QY 848 PYVELKSDGPRDVEVAEAWDNHLNRNRSIVVDLPHGQLRSQVKCKTCGHSVRFPDPFN 907
Db 143 -----RQSDPTP-----QDTSPHDFGGLWRSQIKCLXAG--TSHTTDP-- 180
QY 908 FLSLPLPMS 917
Db 181 FLDVPLDISS 190
RESULT 13
US-09-019-095A-8
; Sequence 8, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; TITLE OF INVENTION: Cell Growth
; FILE REFERENCE: DFCI-435p2A2
; CURRENT APPLICATION NUMBER: US/09/019,095A
; CURRENT FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Murine
US-09-019-095A-8

Query Match 2.3%; Score 200; DB 3; Length 508;
Best Local Similarity 27.8%; Pred. No. 2.1e-08;
Matches 64; Conservative 34; Mismatches 76; Indels 56; Gaps 9;
QY 693 QDECHLVIEVR---NKDMSWPEMSFIANSKIDRHKVPTEKGATGLSLNLTGTCFNMSSI 749
Db 13 QDEAQQVVELTANDEKSLW-----ESQGPQC-GLQNTGNSCYLNAAL 55
QY 750 QCVSNTQPLTQYFISGRHLYELNRT--NPIGMKGMKCYGDLVQELWSGTQKNVAPLKL 807
Db 56 QCLTHPTPLADYMLSOE-----SQCCSPGCK--LCAMEALVTQSLHSHSGDVMKPSH 109
QY 808 RWTIAKYAPRFNGFQQDSQELAFLLDGLHEDLNVRVHEKPYVELKSDGPRDVEVAEAA 867
Db 110 ILTSA-----FKHQQEDAHEFLMFTLETWHESCLQVHRSQETSDES----- 152
QY 868 WDNHLNRNRSIVVDLPHGQLRSQVKCKTCGHSVRFPDPFNFLSLPLPMS 917
Db 153 -----SPIHDFGGLWRSQIKCLCOGTSQSDTYD--RFLDPLDISS 191

RESULT 14

US-09-019-095A-22
; Sequence 22, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; TITLE OF INVENTION: Cell Growth

FILE REFERENCE: DFCI-435p2a2
CURRENT APPLICATION NUMBER: US/09/019,095A
PRIOR FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US96/12884
PRIOR FILING DATE: 1996-08-07
PRIOR APPLICATION NUMBER: US 60/002,066
PRIOR FILING DATE: 1995-08-09
PRIOR APPLICATION NUMBER: US 60/019,787
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 22
LENGTH: 521
TYPE: PRT
ORGANISM: Murine
US-09-019-095A-22

Query Match 2.3%; Score 200; DB 3; Length 521;
Best Local Similarity 27.8%; Pred. No. 2.2e-08;
Matches 64; Conservative 34; Mismatches 76; Indels 56; Gaps 9;
QY 693 QDEQHLVIEVR---NKDMSWPEMSFIANSKIDRHKVPTEKGATGLSLNLTGTCFMNSSI 749
DB 19 QDEAQVVEELTVNGKHSLSW-----ESPQGPC-GLQNTGNSCYLNAAL 61
QY 750 QCVSNTQPLTQYFISGRHLYELNRT--NPIGMKGHMAKCYGDLVQELWSGTQKNVAPLKL 807
DB 62 QCLTHTPPLADYMLSQEH-----SQTCCSPGCK--LCAMEALVTQSLHSHSGDVMPKPSH 115
QY 808 RWTIAKYAPRFNGFQQODSOELLAFLLDGLHEDLNVRVHEKPYVELKDSGRPDWEVAAEA 867
DB 116 ILTSA-----FKHQQEDAHEFLMFTLETWHESCLQVHRQSKPTSEDS-----158
QY 868 WDNHLRRNRSIVVDLPHGQLRSQVKCTCGHISVRFPDFNPLSLPLPMD 917
DB 159 -----SPIHDIRFGWWRWSQIKCLLCQGTSDTYD--RFLDIPLDISS 197

RESULT 15
US-09-019-095A-2
Sequence 2, Application US/09019095A
Patent No. 6287858
GENERAL INFORMATION:
APPLICANT: D'Andrea, Alan D.
APPLICANT: Zhu, Yuan
TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
TITLE OF INVENTION: Cell Growth
FILE REFERENCE: DFCI-435p2a2
CURRENT APPLICATION NUMBER: US/09/019,095A
CURRENT FILING DATE: 1998-02-05
PRIOR FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US96/12884
PRIOR FILING DATE: 1996-08-07
PRIOR APPLICATION NUMBER: US 60/002,066
PRIOR FILING DATE: 1995-08-09
PRIOR APPLICATION NUMBER: US 60/019,787
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 526
TYPE: PRT
ORGANISM: Murine
US-09-019-095A-2

Query Match 2.3%; Score 200; DB 3; Length 526;
Best Local Similarity 27.8%; Pred. No. 2.3e-08;
Matches 64; Conservative 34; Mismatches 76; Indels 56; Gaps 9;
QY 693 QDEQHLVIEVR---NKDMSWPEMSFIANSKIDRHKVPTEKGATGLSLNLTGTCFMNSSI 749
DB 24 QDEAQVVEELTVNGKHSLSW-----ESPQGPC-GLQNTGNSCYLNAAL 66
QY 750 QCVSNTQPLTQYFISGRHLYELNRT--NPIGMKGHMAKCYGDLVQELWSGTQKNVAPLKL 807

DB 67 QCLTHTPPLADYMLSQEH-----SQTCCSPGCK--LCAMEALVTQSLHSHSGDVMPKPSH 120
QY 808 RWTIAKYAPRFNGFQQODSOELLAFLLDGLHEDLNVRVHEKPYVELKDSGRPDWEVAAEA 867
DB 121 ILTSA-----FKHQQEDAHEFLMFTLETWHESCLQVHRQSKPTSEDS-----163
QY 868 WDNHLRRNRSIVVDLPHGQLRSQVKCTCGHISVRFPDFNPLSLPLPMD 917
DB 164 -----SPIHDIRFGWWRWSQIKCLLCQGTSDTYD--RFLDIPLDISS 202

Search completed: February 11, 2004, 15:32:25
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:30:16 ; Search time 45 Seconds
(without alignments)
7463.315 Million cell updates/sec

Title: US-09-888-615-73

Perfect score: 8563

Sequence: 1 MGAKESRIGFLSYVEALRRV.....DTSSMDEDFSDYKVCVLQ 1604

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTU5_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8563	100.0	1604	9	US-09-888-615-73
2	6826	79.7	1317	12	US-10-071-275-2
3	4428	51.7	847	12	US-10-094-749-3045
4	3138	36.6	582	12	US-10-274-639-14
5	2711	31.7	539	12	US-09-764-875-762
6	1750	20.4	401	12	US-10-158-057-222
7	1406	16.4	786	15	US-10-071-838-17
8	1124	13.1	1041	12	US-10-369-493-5738
9	994	11.6	827	12	US-10-210-130-84
10	890	10.4	173	11	US-09-764-881-153
11	890	10.4	173	12	US-10-243-747-153
12	890	10.4	173	12	US-10-158-057-300
13	890	10.4	173	12	US-09-764-875-1064
14	787.5	9.2	973	12	US-09-493-22797
15	756.5	8.8	900	12	US-10-369-493-6013

16	738.5	8.6	1254	12	US-10-369-493-22131	Sequence 22131, A
17	675	7.9	1095	12	US-10-369-493-5583	Sequence 5583, Ap
18	665	7.8	924	12	US-10-369-493-22682	Sequence 22682, A
19	665	7.8	1318	12	US-10-291-172-338	Sequence 338, App
20	646.5	7.5	929	12	US-10-369-493-3912	Sequence 3912, Ap
21	530.5	6.2	680	12	US-10-369-493-12503	Sequence 12503, A
22	529	6.2	292	14	US-10-105-992-5	Sequence 5, Appl
23	524	6.1	1123	9	US-09-888-615-77	Sequence 77, Appl
24	498	5.8	441	15	US-10-153-668-102	Sequence 102, App
25	498	5.8	634	15	US-10-153-668-98	Sequence 98, Appl
26	498	5.8	634	15	US-10-153-668-100	Sequence 100, App
27	498	5.8	1118	15	US-10-153-668-104	Sequence 104, Appl
28	488	5.7	1080	12	US-10-205-219-62	Sequence 62, Appl
29	408	4.8	605	12	US-10-094-749-2994	Sequence 2994, Ap
30	371	4.3	926	12	US-10-369-493-1623	Sequence 1623, Ap
31	330	3.9	805	12	US-10-369-493-1769	Sequence 1769, Ap
32	321	3.7	565	10	US-09-766-954A-2	Sequence 2, Appl
33	318.5	3.7	489	12	US-09-833-245-211	Sequence 211, App
34	314	3.7	463	12	US-09-833-245-212	Sequence 212, App
35	288	3.4	435	12	US-10-264-237-2761	Sequence 2761, Ap
36	280	3.3	814	9	US-09-808-568-5	Sequence 5, Appl
37	280	3.3	814	12	US-10-098-108-7	Sequence 7, Appl
38	280	3.3	814	15	US-10-156-239-52	Sequence 52, Appl
39	273.5	3.2	812	15	US-10-097-340-330	Sequence 330, App
40	271	3.2	593	9	US-09-872-153-18	Sequence 18, Appl
41	257.5	3.0	207	12	US-10-264-049-4055	Sequence 4055, Ap
42	257.5	3.0	953	9	US-09-888-615-66	Sequence 66, Appl
43	249.5	2.9	1071	12	US-10-369-493-22077	Sequence 22077, A
44	249	2.9	656	12	US-10-274-639-20	Sequence 20, Appl
45	237.5	2.8	783	9	US-09-888-615-67	Sequence 67, Appl

ALIGNMENTS

RESULT 1

US-09-888-615-73
; Sequence 73, Application US/0988615
; Patent No. US20020044858A
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARVDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 73
; LENGTH: 1604
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-73

Query Match	100.0%;	Score 8563;	DB 9;	Length 1604;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1604;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGAKESRIGFLSYVEALRRVTDVELKELKDAFKRTCTGLSYMGCHCFIRVLGDDGVPVKY	60	
Db	1	MGAKESRIGFLSYVEALRRVTDVELKELKDAFKRTCTGLSYMGCHCFIRVLGDDGVPVKY	60	
Qy	61	AEVITYCSFGTSGKLHNNLIIVGLVLLTRCKDEKAKYIFSLFSSSGNVIREEVERML	120	
Db	61	AEVITYCSFGTSGKLHNNLIIVGLVLLTRCKDEKAKYIFSLFSSSGNVIREEVERML	120	
Qy	121	HVDDGKVPDTRLCFSGSEKKNYKFNWFLNKNDAFTFSRWLLSGGVYVTLTDDSDTPT	180	

Db 121 HVVDGKVPDLKCRSEGEKQYEFKRWLFLNKDAFTSRMLLSGGVYVLTDDSDPT 180
 QY 181 FYQTLAGVTHLESDDIIDEKRYWLLKAQSRGTGRDLETFGLVPSPPIRPSLSSEGLFNAF 240
 Db 181 FYQTLAGVTHLESDDIIDEKRYWLLKAQSRGTGRDLETFGLVPSPPIRPSLSSEGLFNAF 240
 QY 241 DENRDNHIDFKISCSLACCGRPLAEROKFCFVDVDRDGLSRVLRDMVVALLEVW 300
 Db 241 DENRDNHIDFKISCSLACCGRPLAEROKFCFVDVDRDGLSRVLRDMVVALLEVW 300
 QY 301 KDNRTDDIPELHMDLSDIYEGILNAHDHDKMGLHLEDTYQIWSVKNVLANFLNLLFOVC 360
 Db 301 KDNRTDDIPELHMDLSDIYEGILNAHDHDKMGLHLEDTYQIWSVKNVLANFLNLLFOVC 360
 QY 361 HVLGLRPAATPEEGQIRGMLERESRYGLQAGHWNFIISQWQWQWKEYKYKNDANPVI 420
 Db 361 HVLGLRPAATPEEGQIRGMLERESRYGLQAGHWNFIISQWQWQWKEYKYKNDANPVI 420
 QY 421 EPSVVLNGGKYSGFTAHPMEQVEDRIGSSLSYVNTTEKPSDNISTASESETAGSGFL 480
 Db 421 EPSVVLNGGKYSGFTAHPMEQVEDRIGSSLSYVNTTEKPSDNISTASESETAGSGFL 480
 QY 481 YSATPGADVCFARQHTSDNNQCLLGANGNILLHLNPKQGAINDQPLVTOEPVKATSL 540
 Db 481 YSATPGADVCFARQHTSDNNQCLLGANGNILLHLNPKQGAINDQPLVTOEPVKATSL 540
 QY 541 TLEGGRLKRTPOLIHRDGYEMVPEPVWRAHYWYGANLALPRVINKNSKTDIPELELPPR 600
 Db 541 TLEGGRLKRTPOLIHRDGYEMVPEPVWRAHYWYGANLALPRVINKNSKTDIPELELPPR 600
 QY 601 YLLFLRQOPATRTQOSNIWVWNGVSPNAPLKRVLATYGCFSRMTQIKEIHEYLSQRLR 660
 Db 601 YLLFLRQOPATRTQOSNIWVWNGVSPNAPLKRVLATYGCFSRMTQIKEIHEYLSQRLR 660
 QY 661 IKEEDMRLWYNSYNTLLDDEHKLKYLKIODEOHLVIEVRNKMSPMSFIANS 720
 Db 661 IKEEDMRLWYNSYNTLLDDEHKLKYLKIODEOHLVIEVRNKMSPMSFIANS 720
 QY 721 KIDRHVPTKEGATGSLNLTGNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRTNPIGMK 780
 Db 721 KIDRHVPTKEGATGSLNLTGNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRTNPIGMK 780
 QY 781 GHMAKCYGDLVQELWSGTQKNVAPLKLRTIAYAPRNGFOQDSQELLAFLLDGLHED 840
 Db 781 GHMAKCYGDLVQELWSGTQKNVAPLKLRTIAYAPRNGFOQDSQELLAFLLDGLHED 840
 QY 841 LNRVHEKPYVELKDSGRPDWEAAEAWNDHLRNRISIVVDLPHGLRQSRQVCKCTCGHIS 900
 Db 841 LNRVHEKPYVELKDSGRPDWEAAEAWNDHLRNRISIVVDLPHGLRQSRQVCKCTCGHIS 900
 QY 901 VRDPPNFVLSPLPMDSYMHLEITVTKLGGTTPVRYGLRLNDEKTYGLKQSLDGLGN 960
 Db 901 VRDPPNFVLSPLPMDSYMHLEITVTKLGGTTPVRYGLRLNDEKTYGLKQSLDGLGN 960
 QY 961 SEQILLAEVHGSNIKNFPDQNKVRLSVSGFLCAFEIPVPSPIASSPTQTDRESSPST 1020
 Db 961 SEQILLAEVHGSNIKNFPDQNKVRLSVSGFLCAFEIPVPSPIASSPTQTDRESSPST 1020
 QY 1021 NEMFTLTNGDLPRPIFIPNGMNTVPPCGTEKNFTNGMVGHPMLSPDPSPTGTIIVH 1080
 Db 1021 NEMFTLTNGDLPRPIFIPNGMNTVPPCGTEKNFTNGMVGHPMLSPDPSPTGTIIVH 1080
 QY 1081 RKMRTLYFLSSQKRPBSLFGMPLVPCVTHTKKDLYDAVWIOVSRLASPLPQEAAS 1140
 Db 1081 RKMRTLYFLSSQKRPBSLFGMPLVPCVTHTKKDLYDAVWIOVSRLASPLPQEAAS 1140
 QY 1141 HAQCDSDSMGYQYFPTLRVYQKDGNSCAMCPWYRRCRGCKIDCGEDRAFIAGNAYIAVDWD 1200
 Db 1141 HAQCDSDSMGYQYFPTLRVYQKDGNSCAMCPWYRRCRGCKIDCGEDRAFIAGNAYIAVDWD 1200
 QY 1201 PTALHLRYQTSQBRVDEHESVQSRQAEPINLDSCLRAFTSBEELGENEMYCCK 1260

Db 1201 PTALHLRYQTSQBRVDEHESVQSRQAEPINLDSCLRAFTSBEELGENEMYCCK 1260
 QY 1261 THCLATKLDLWRLPILLIHLKRFQFVNGRWIKSOKIVKPPRESPPDPSAFLVPRDPALC 1320
 Db 1261 THCLATKLDLWRLPILLIHLKRFQFVNGRWIKSOKIVKPPRESPPDPSAFLVPRDPALC 1320
 QY 1321 QHKLPLTPQGDDELSEPRILAREVKVDAQSSAGBEDVLLSKSPSSISANIISSPKGSPSS 1380
 Db 1321 QHKLPLTPQGDDELSEPRILAREVKVDAQSSAGBEDVLLSKSPSSISANIISSPKGSPSS 1380
 QY 1381 RKSGTSCPSKNSPNSPRTLGRSKRLRLPQIGSKNKLSSKENLDAKNGAGQICE 1440
 Db 1381 RKSGTSCPSKNSPNSPRTLGRSKRLRLPQIGSKNKLSSKENLDAKNGAGQICE 1440
 QY 1441 LADALSRGHVILGGSQBELVTPDQHEVALANGFLYEACNGVSNQGLGNHSEEDSTDQ 1500
 Db 1441 LADALSRGHVILGGSQBELVTPDQHEVALANGFLYEACNGVSNQGLGNHSEEDSTDQ 1500
 QY 1501 REDTRIKPIYNLVYAISCHSGILGGGHYVTVYAKNPCKWICYNDSSCKELHPDIDTDSAY 1560
 Db 1501 REDTRIKPIYNLVYAISCHSGILGGGHYVTVYAKNPCKWICYNDSSCKELHPDIDTDSAY 1560
 QY 1561 ILFYEQGGIDYAOFLPKTDGKWADTSSMDEDPESDYKKYCVLQ 1604
 Db 1561 ILFYEQGGIDYAOFLPKTDGKWADTSSMDEDPESDYKKYCVLQ 1604
 RESULT 2
 US-10-071-275-2
 ; Sequence 2, Application US/10071275
 ; Publication No. US20030138934A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Meyers, Rachel E.
 ; TITLE OF INVENTION: 80091, A NOVEL HUMAN UBIQUITIN
 ; FILE OF INVENTION: CARBOXY-TERMINAL HYDROLASE FAMILY MEMBER AND USES THEREOF
 ; TITLE REFERENCE: 10448-139001
 ; CURRENT APPLICATION NUMBER: US/10/071,275
 ; PRIOR FILING DATE: 2002-02-07
 ; PRIOR FILING DATE: 2001-02-07
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 1317
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-071-275-2
 Query Match 79.7%; Score 6826; DB 12; Length 1317;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1274; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 330 KMGHLTLEDYQIWSVKNVLANFLNLLFQVCHIVLGLRPAATPEEGQIRGMLERESRYG 389
 Db 43 QMGHLTLEDYQIWSVKNVLANFLNLLFQVCHIVLGLRPAATPEEGQIRGMLERESRYG 102
 QY 390 LQAGHWNFIISQWQWQWKEYKYKNDANPVIPESSVNLNGKYSFGTAHPMEQVEDRIGS 449
 Db 103 LQAGHWNFIISQWQWQWKEYKYKNDANPVIPESSVNLNGKYSFGTAHPMEQVEDRIGS 162
 QY 450 SLSVYNTTEKFSNDISTASESETAGSGFLYATPGADVCFARQHTSDNNQCLLGAN 509
 Db 163 SLSVYNTTEKFSNDISTASESETAGSGFLYATPGADVCFARQHTSDNNQCLLGAN 222
 QY 510 GNILLHLNPKQGAINDQPLVTOEPVKATSLTLEGGRLKRTPOLIHRDGYEMVPEPVWRA 569
 Db 223 GNILLHLNPKQGAINDQPLVTOEPVKATSLTLEGGRLKRTPOLIHRDGYEMVPEPVWRA 282
 QY 570 LYHWYGANLALPRVINKNSKTDIPELELFPYLLFLRQOPATRTQOSNIWVWNGVSPN 629
 Db 283 LYHWYGANLALPRVINKNSKTDIPELELFPYLLFLRQOPATRTQOSNIWVWNGVSPN 342
 QY 630 APLKRVLATYGCFSRMTQIKEIHEYLSQRLR I KEEDMRLWYNSYNTLLDDEHKLKYL

Db 343 APLKRVLAYTCFERNQTIKEIHLYLQRLKEEDNRLWLNSENVLTLDDDEHKEY 402
Qy 690 LKIQDEHLVTEVNKONKSWPEMSFIANGSKIDRHVKPTKAGATGSLNNTCFMNSSI 749
Db 403 LKIQDEHLVTEVNKONKSWPEMSFIANGSKIDRHVKPTKAGATGSLNNTCFMNSSI 462
Qy 750 QCVNTPLOTOYFISGRHLYELNNTPIGMKGMAKCYGDLVQELKSGTQKNVAPLKLW 809
Db 463 QCVNTPLOTOYFISGRHLYELNNTPIGMKGMAKCYGDLVQELKSGTQKNVAPLKLW 522
Qy 810 TIAYAPRFNGFOQDSQELLAFLLDGLHEDLNKVEKPYVELKSDGRPDWEAAEAWD 869
Db 523 TIAYAPRFNGFOQDSQELLAFLLDGLHEDLNKVEKPYVELKSDGRPDWEAAEAWD 582
Qy 870 NHLRNSIVVDLPHGOLRSOVCKTGGHSVREDPENFLSLPLPMDSYNHLBITVIKLD 929
Db 583 NHLRNSIVVDLPHGOLRSOVCKTGGHSVREDPENFLSLPLPMDSYNHLBITVIKLD 642
Qy 930 GTTPVRYGLRLNMDCKYTGKQLSDLCGLNSQIILLAEVHGNIKNFPDQNKVRLSVS 989
Db 643 GTTPVRYGLRLNMDCKYTGKQLSDLCGLNSQIILLAEVHGNIKNFPDQNKVRLSVS 702
Qy 990 GFLCAPFIPVPSISASPTQTFSSPSSTNEMFTLTNGDLPRPIFIPNGMNTVWPC 1049
Db 703 GFLCAPFIPVPSISASPTQTFSSPSSTNEMFTLTNGDLPRPIFIPNGMNTVWPC 762
Qy 1050 GTEKNFTNGMNGHMPISLPSPFTGYIIIAVRKMRTELFLSOKNRPDLFGWPLIVPC 1109
Db 763 GTEKNFTNGMNGHMPISLPSPFTGYIIIAVRKMRTELFLSOKNRPDLFGWPLIVPC 822
Qy 1110 TVHTRKDLDAVMIQVSRLASPLPPOEASNAHQDSDSGYQYFPLRVVQDGNSCAW 1169
Db 823 TVHTRKDLDAVMIQVSRLASPLPPOEASNAHQDSDSGYQYFPLRVVQDGNSCAW 882
Qy 1170 CPWFRGCGCKIDCEBRAFTGNAYIAVDWDPALHLYRQTSQERVVDEHSEVQSRRAQ 1229
Db 883 CPWFRGCGCKIDCEBRAFTGNAYIAVDWDPALHLYRQTSQERVVDEHSEVQSRRAQ 942
Qy 1230 AEPINLDSCLRAFTSEBELGENEMYCKCKTHCLATKCLDLWLRLPILIIHLKRFQFVN 1289
Db 943 AEPINLDSCLRAFTSEBELGENEMYCKCKTHCLATKCLDLWLRLPILIIHLKRFQFVN 1002
Qy 1290 GRWIKSQIKVFPRESFDPSPALVPRDPALCOHKLTPQGDSELPRIAREVKVQDAQS 1349
Db 1003 GRWIKSQIKVFPRESFDPSPALVPRDPALCOHKLTPQGDSELPRIAREVKVQDAQS 1062
Qy 1350 SAGEDVLLSKSPSLSANIISPKGSPSSRSGTSCPSKSSPNSSPRTILGRSKGRL 1409
Db 1063 SAGEDVLLSKSPSLSANIISPKGSPSSRSGTSCPSKSSPNSSPRTILGRSKGRL 1122
Qy 1410 RLPQIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLGGSQPELTPQDHEVALA 1469
Db 1123 RLPQIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLGGSQPELTPQDHEVALA 1182
Qy 1470 NGELYEHACNGYNGOLGNHSEEDTDQRETRIKPIYNLYASCHSGIIGGGHYVT 1529
Db 1183 NGELYEHACNGYNGOLGNHSEEDTDQRETRIKPIYNLYASCHSGIIGGGHYVT 1242
Qy 1530 YAKPNCKWYCYNDSSCKELHPDEIDTDSAYILFYEQQIDYAOFLPKTDGKMAUTSSM 1589
Db 1243 YAKPNCKWYCYNDSSCKELHPDEIDTDSAYILFYEQQIDYAOFLPKTDGKMAUTSSM 1302
Qy 1590 DEDFESDYKCYVLQ 1604
Db 1303 DEDFESDYKCYVLQ 1317

RESULT 3

US-10-094-749-3045

; Sequence 3045, Application US/10094749

; Publication No. US20030219741A1

; GENERAL INFORMATION:

APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOTHIKO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOKUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 3045
LENGTH: 847
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-3045

Query Match 51.7%; Score 4428; DB 12; Length 847;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 831; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 19 RYTDVELKELKDAFRTCGLSYMGQHCFFIREVLGDGVPPKVAEVIYCSFGGTSGKLHFN 78
Db 16 RYTDVELKELKDAFRTCGLSYMGQHCFFIREVLGDGVPPKVAEVIYCSFGGTSGKLHFN 75
Qy 79 NLIVGLVLLTRGKDEKAKYIFSLFSSSESGNYVIREEMERMLHVVDGKVPDTRKCFSEG 138
Db 76 NLIVGLVLLTRGKDEKAKYIFSLFSSSESGNYVIREEMERMLHVVDGKVPDTRKCFSEG 135
Qy 139 EKVNYEKEFERNLFLNKDAFTFSRMLLSGVVYVTLTDDSDTPTFYQTLAGVTHLESDDI 198
Db 136 EKVNYEKEFERNLFLNKDAFTFSRMLLSGVVYVTLTDDSDTPTFYQTLAGVTHLESDDI 195
Qy 199 LEKRYWLLKAQSRTRGRFPLETFGPIVSPPIRPSISEGLFNAPDENRDNHIDFKETSCGLS 258
Db 196 LEKRYWLLKAQSRTRGRFPLETFGPIVSPPIRPSISEGLFNAPDENRDNHIDFKETSCGLS 255
Qy 259 ACCRPLAEROKCFKPFVDVDRDGVLSRVELDMVVALLEYVKONRTDDIPELHMDLSDI 318
Db 256 ACCRPLAEROKCFKPFVDVDRDGVLSRVELDMVVALLEYVKONRTDDIPELHMDLSDI 315
Qy 319 VEGILNAHDTTKMGHLLTLEDYQIWSVKVNLANEFLNLIFQVCHIVGLRPARPPEEGQII 378
Db 316 VEGILNAHDTTKMGHLLTLEDYQIWSVKVNLANEFLNLIFQVCHIVGLRPARPPEEGQII 375
Qy 379 RGWLERESRYGLQAGHNFIIISMQWQWKYKYKADNPVIEPSSVLNGGKYSFGTAAH 438
Db 376 RGWLERESRYGLQAGHNFIIISMQWQWKYKYKADNPVIEPSSVLNGGKYSFGTAAH 435
Qy 439 PMEQVEDRIGSSLSYVNTTEKFSNISTASEASETAGSFLYSATPGADVCFARQHNIS 498
Db 436 PMEQVEDRIGSSLSYVNTTEKFSNISTASEASETAGSFLYSATPGADVCFARQHNIS 495
Qy 499 DNNOCILGANGNILLHLNLPKQPGAINQPLVTPQEPVKATSLITLEGGRKKTPTQLIHGRD 558
Db 496 DNNOCILGANGNILLHLNLPKQPGAINQPLVTPQEPVKATSLITLEGGRKKTPTQLIHGRD 555

QY 559 YEMPEPVWVRLHYHYGANLALPRVVKNSKTDIPELELFPYLLFLRQOPATRTQGSNI 618
Db 556 YEMPEPVWVRLHYHYGANLALPRVVKNSKTDIPELELFPYLLFLRQOPATRTQGSNI 615
QY 619 WYNGNVPSPNAPLKRVLAYTGCFSRMOTIKIEIHEYLSORLRIKEEDMRLWLYNSENYLT 678
Db 616 WYNGNVPSPNAPLKRVLAYTGCFSRMOTIKIEIHEYLSORLRIKEEDMRLWLYNSENYLT 675
QY 679 LDDDEHKLLEYLKIODEQHLVIEVRNKMSPPEMSFIANSKIDRHKVPTEKATGLSN 738
Db 676 LDDDEHKLLEYLKIODEQHLVIEVRNKMSPPEMSFIANSKIDRHEVPTKATGLSN 735
QY 739 LGNTCFMSSIOCVSNTQPTQYFISGRHLYELNETNPIGMKGHWAKCYGDLVQELWSGT 798
Db 736 LGNTCFMSSIOCVSNTQPTQYFISGRHLYELNETNPIGMKGHWAKCYGDLVQELWSGT 795
QY 799 QXNVAPLKLRTWITAKYAPRFNGFOQDQSQELLAFLLDGLHEDLNRVHEKPYV 850
Db 796 QXNVAPLKLRTWITAKYAPRFNGFOQDQSQELLAFLLDGLHEDLNRVHEKPYV 847

RESULT 4

US-10-274-639-14
; Sequence 14, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DELSEANE, Angelo M.; GANDHI, Ameena R.
; APPLICANT: RAFALIA, April J.A.; LU, Dying Aina M.
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.
; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.
; APPLICANT: NGUYEN, Darniel B.; LEE, Ernestine A.
; APPLICANT: KHAN, Farrah A.; YUE, Henry
; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
; APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha
; APPLICANT: DING, Li; KEARNEY, Liam
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.
; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.
; APPLICANT: BURFORD, Neil; WALIA, Narinder K.
; APPLICANT: LAL, Preeti G.; LEE, Sally
; APPLICANT: TODD, Stephen; LO, Terence P.
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
; APPLICANT: AZIMZAI, Yalda; LU, Yan
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0167 USA
; CURRENT APPLICATION NUMBER: US/10/274,639
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: PCT/US01/22397
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/220,063
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/221,680
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/223,544
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/224,717
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/225,988
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 60/227,568
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 14
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030232349A1 7476667CD1
US-10-274-639-14

Query Match 36.6%; Score 3138; DB 12; Length 582;
Best Local Similarity 100.0%; Pred. No. 9.5e-253;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1023 MFTLTITNGDLPPIPIFIPNGMNTVVPFCGTEKXFTNGMVGHPSPDPGPFYTCYIIIVHRK 1082
Db 1 MFTLTITNGDLPPIPIFIPNGMNTVVPFCGTEKXFTNGMVGHPSPDPGPFYTCYIIIVHRK 60
QY 1083 MMRTELYFSSQKRPSPFPGMPLIVPCTVHTTKKDYDAVMTQVSRLASPLPPOASNAH 1142
Db 61 MMRTELYFSSQKRPSPFPGMPLIVPCTVHTTKKDYDAVMTQVSRLASPLPPOASNAH 120
QY 1143 QDCDDSMGYQYFPTLVVQKDGNSCAWCFWYFCGCKIDCGEDRAFIAGNIAIVDWDPT 1202
Db 121 QDCDDSMGYQYFPTLVVQKDGNSCAWCFWYFCGCKIDCGEDRAFIAGNIAIVDWDPT 180
QY 1203 ALHLRYQTSQERVVDHESVESQSRRAQAPINLDSCLRAFTSEELGENEMYCCKKTH 1262
Db 181 ALHLRYQTSQERVVDHESVESQSRRAQAPINLDSCLRAFTSEELGENEMYCCKKTH 240
QY 1263 CLATKKLDLWRLPPLIILHLKSPQFVNGEWIKSQIVKFPRESFDPSPAFVPRDPALCOH 1322
Db 241 CLATKKLDLWRLPPLIILHLKSPQFVNGEWIKSQIVKFPRESFDPSPAFVPRDPALCOH 300
QY 1323 KPILTPQDSELSPPRIILAREVVKVDAQSSAGEEDVLLSKSPSSLSANIISPPKSPSSSRK 1382
Db 301 KPILTPQDSELSPPRIILAREVVKVDAQSSAGEEDVLLSKSPSSLSANIISPPKSPSSSRK 360
QY 1383 SGTSCPSSKXNSPNSPRTLGSRKGLRLPQIGKSKLSSKENLDASKENGAGQICELA 1442
Db 361 SGTSCPSSKXNSPNSPRTLGSRKGLRLPQIGKSKLSSKENLDASKENGAGQICELA 420
QY 1443 DALSRGHVJGSGQPELVTPQDHEVALANGFLYEACGNGYNGQLGHNSEEDSTDDQRE 1502
Db 421 DALSRGHVJGSGQPELVTPQDHEVALANGFLYEACGNGYNGQLGHNSEEDSTDDQRE 480
QY 1503 DTRIKPIYMLYAISSCHSGLGGHYVYAKNPCKWYCYNDSSCKELHHPDEIDTDSAYIL 1562
Db 481 DTRIKPIYMLYAISSCHSGLGGHYVYAKNPCKWYCYNDSSCKELHHPDEIDTDSAYIL 540
QY 1563 FYEQQGDYVAQFLPKTDGKKMADTSSMDEDFESDYKCYVLQ 1604
Db 541 FYEQQGDYVAQFLPKTDGKKMADTSSMDEDFESDYKCYVLQ 582

RESULT 5
US-09-764-875-762
; Sequence 762, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 762
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-875-762

Query Match 31.7%; Score 2711; DB 12; Length 539;
Best Local Similarity 93.3%; Pred. No. 4e-217;
Matches 512; Conservative 9; Mismatches 6; Indels 22; Gaps 2;
QY 117 ERMHLVVVDGKVPDTRKCFSEGEKYNVEKFRNWLFLNKDAFTSRWLLSGGVVYITDSD 176
Db 7 ERMHLVVVDGKVPDTRKCFSEGEKYNVEKFRNWLFLNKDAFTSRWLLSGGVVYITDSD 66

ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human TRE-2/USP6
US-10-071-838-17

Query Match 16.4%; Score 1406; DB 15; Length 786;
Best Local Similarity 92.7%; Pred. No. 8.5e-108;
Matches 266; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 701 EVRNKMSWPEEMSPFANSKIDRHKVPTEKGATGLSNLGNCTCFMNSSIOCVSNTOPLTQ 760
DB 499 EVHNKMSWPEEMSPFANSKIDRHKVPTEKGATGLSNLGNCTCFMNSSIOCVSNTOPLTQ 558
QY 761 YFISGRHLYELNRNTPIGMKGMKACYGDLYQELWSGTQKNVAPLKLRTWITAKYAPFNG 820
DB 559 YFISGRHLYELNRNTPIGMKGMKACYGDLYQELWSGTQKNVAPLKLRTWITAKYAPFNG 618
QY 821 FQOQDSQELLAFLLDGLHEDLNVRHEKPYVELKSDGRPDWEVAABAWDNHLNRNRIIV 880
DB 619 FQOQDSQELLAFLLDGLHEDLNVRHEKPYVELKSDGRPDWEVAABAWDNHLNRNRIIV 678
QY 881 DLFHQLSQVKKCTCGHISVRFPFNFSLPLPMDSYMHLEITVVKLDGTPVRYGLRL 940
DB 679 DLFHQLSQVKKCTCGHISVRFPFNFSLPLPMDSYMHLEITVVKLDGTPVRYGLRL 738
QY 941 NMDEKYTGKQLSDLCGLNSEQILLAEVHGSNINFPQDNQKVRLS 987
DB 739 NMDEKYTGKQLSDLCGLNSEQILLAEVHGSNINFPQDNQKVRLS 785

RESULT 8

US-10-369-493-5738

Sequence 5738, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 5738

LENGTH: 1041

TYPE: PRT

ORGANISM: Caenorhabditis elegans

US-10-369-493-5738

Query Match 13.1%; Score 1124; DB 12; Length 1041;
Best Local Similarity 24.3%; Pred. No. 5.1e-84;
Matches 358; Conservative 188; Mismatches 436; Indels 490; Gaps 40;

QY 141 VNYEKRNLFLNKDAFTSRMLSGGVVYVTLTDDSDTPTFYQTLAGVTHLEESDIIDLE 200
DB 16 IGAEADAKTYRFI-EYHFEASFAGLSPKLTIPLNSYI-SFYEVMSYVTHLSVNEVIELE 73
QY 201 KRYWLLKAQSRGTFPLETFGLVPPPIRPSLSEGLFNAFENDRNHDHFKIEISGLSAC 260
DB 74 KVPATISDRA-VCKLNEEKWQALGCGFPDSYAERLFAVFDENRQDIDFRELVCITLSAL 132
QY 261 CRGFLAERQKFKFVDVDRDGVLSRVELRDMVALLEWKNRDTDDIPELMOLSDIVE 320
DB 133 CRGFLPGRISQALRIWVDVDCDKLLSDELSNM-----YKDL---NVPEEHQVTV--- 178
QY 321 GILNAHDTTKWGHILTLEDYQIWSVKM-VLANEFLLNLFQVCHIVLGLRPAATPEEGQIIR 379
DB 179 -----KSSNGKSALVDGFIWAQENEKYVNEYYSVALQIGHICLGLAPESRKMLQIWN 231

QY 380 GWLEBESRYGLQAGHNWELISMQWQOQKEYYKYDANPNVWIEPSSVLNGKYSFCTAAHP 439
DB 232 EFEKASELPLS---EWNIVASGWAELRSLEADKNPNIDNSOI----- 274
QY 440 MBQVEDRIGSSLSYNTTTEKFSNISTASEASETAGSGFLYSATPGADVCFARQHTSD 499
DB 275 -----KGTREDSWTS----- 284
QY 500 NNNQCLLGANGNILLHNPQKGAIDNQPLYTQEPVKATSLTLEGRLKRPQLLHGRDY 559
DB 285 -----KVACISAESARLK---PDLIPS-DY 305
QY 560 EMVPPVWREALYHYTCANLALPRPVKN-----SKTDIPELELPYRYLLFL---RQOPA 610
DB 306 IRVPVPLWEALRWHGCAITVDSQFTRKYLDGEFFDNKPALEYLPLEILLGHDRKKSQ 365
QY 611 TTIQOSNIWNMGVNPSPNAPLKRVLAYTGC---FSRMTIKIEIHEIYLSQRRIKEDARL 668
DB 366 DQTE-----NTP---RSLTSMWACQVSRSMVTVDLALCKTELRLGDDARL 409
QY 669 WLYNSENY-LTLDDDEHKLLEYL-----KIQDEQHLVIEVRNKDMS-WPEEM----- 713
DB 410 WQVVKENEENGVLLDDGAQNLHQLYSSLGKTKKVKMKLLLEVRERGTVGWEELRASLS 469
QY 714 --SFIANSKIDRHKVPTEKGATGLSNLGNCTCFMNSSIOCVSNTOPLTQYFISGRHLVEL 771
DB 470 GKQITAAASTLSSNAQLSGRPGAVGLVNYGNFYRAGIQCLARVSPLTQYFLDEDNLDAL 529
QY 772 NRTNPIGMKGMKAC---YGDLYQELWSGTQKNVAPLKLRTWITAKYAPFNGFQOQDSOE 828
DB 530 KRGNL---RGGDAETTIEYAKLLRENWAKKNIAF----- 563
QY 829 LLAFLDLGHLHEDLNVRHEKPYVELKSD--GRPDWEVAABAWDNHLNRNRIIVDLFHGOL 887
DB 564 -----NDL-----IKKEADKEADEKAERSWTEYEKQNESLVTQFTGOL 605
QY 888 RSQVCKCTCGHISVRFPFNFSLPLPMDSYMHLEITVVKLDGTPVRYGLNLDKVT 947
DB 606 RSKLICRTQSSSSVPEPTSLPIGDFEDVLYQYVVRDGRIPRRYIGFLSRDSKVG 665
QY 948 GLKKQLSDLCGLNSEQILLAEVHGSNINFPQDNQKVRLSVSGFLCAFEIPVPVPSIPAS 1007
DB 666 NLREVVAVSSGISMSHLTIQ-----CMSSKGTL-----MSRSPNERS 702
QY 1008 SPTQTD--FSSSPSTNEMFTLTNGDLPRPIFIPNGMPTVVPCTGTEKNFTNGMNGHP 1065
DB 703 SNLRDELPLSSFPFGARLYAL----- 723
QY 1066 SLPDSPTGT---YIIAVHRKMM-RTELYFLSSQKN-RPSLFQMPLIVPCTVHTRKKDLYD 1120
DB 724 ELPESS--TGEDQWRVAMERKLQYNHEPYILGSTAGFIVSRFGLPLIVGLDEEVTGKLYE 781
QY 1121 AVMIQVSRLASPLFPOEASNAHQDQDDSMGYQYPTFLRVVQKDGNSCAWCPYRRCRGCK 1180
DB 782 DVMIQVHREHESVNSSSRRAHDPCDE--NSGYPTTCLVDPNYEWCGCPALRCRGP 840
QY 1181 IDCGEDRAFT-GNAYTAVDWDPTALHLYCTQERVDEHESVRSRAQAAPINLDSCL 1239
DB 841 IRDESKVIFPANCPIAVDWLFIYLRVYNHQBQACEDDPSVAETWSRHFAPSSLEHCL 900
QY 1240 RAFTSEELGENEMYCSKCTHCLATKKLDLWRJPPILIIHLKRFQV--NGRWIKSQK 1297
DB 901 EKSPCEPL--DAAIQDCRCKTMDKVTWKLPKYLIIHLKRFEFLEQGRMGKCKR 958
QY 1298 IVKFPRESDDPSAFVPRDPALCQHKPLTPQGDSELPILAREVKVKVDAQSSAGEDVL 1357
DB 959 TVNFPKJGHFDPAFF----- 972
QY 1358 LSKSPSLSANITSSPKGSPSSSRKSGTSCPSKNSPNSPRTLGRSKGRRLPQIGSK 1417
DB 973 -----VDKPDGN----- 979

QY 1418 NKLSSKENLDASKENGAGQICELADALSRGHVLGSGQPELVTPQDHEVALANGFLYEHE 1477
Db 980 ----- 979
QY 1478 ACGNGYSNGQLGNHSEEDSTDQREDTRIKPIYNYIAISCHSGLGGHYTYAKNPCK 1537
Db 980 -----TYECIALANHYQLSCGHFYAKSNEBK 1008
QY 1538 WYCYNDSSCKELHPDEIDTDSAYILFYEQGI 1569
Db 1009 WLLNDSCVRESEEVXQGYLLFYERKDV 1040

RESULT 9

US-10-210-130-84
; Sequence 84, Application US/10210130
; Publication No. US20040014053A1

GENERAL INFORMATION:

; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Patturajan, Meera
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Miller, Charles E.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Li, Li
; APPLICANT: Berghs, Constance
; APPLICANT: Zhong, Mei
; APPLICANT: Casman, Stacie J.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Smithson, Glennda
; APPLICANT: Ji, Weizhen
; APPLICANT: Gorman, Linda
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Leite, Mario W.
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Anderson, David W.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Khramtsov, Nikolai V.
; APPLICANT: Ort, Tatiana
; APPLICANT: Ellerman, Karen
; APPLICANT: Rastelli, Luca
; APPLICANT: Agee, Michele L.
; APPLICANT: Chaudhuri, Amitabha
; APPLICANT: Chant, John S.
; APPLICANT: DiPippo, Vincent A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Gangolli, Bsha A.
; APPLICANT: Giot, Loic
; APPLICANT: Ooi, Chean Eng
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Spaderina, Steven K.
; APPLICANT: Hjalt, Tord
; APPLICANT: Liu, Xiaohong
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Catterton, Elna
; APPLICANT: Shenoy, Suresh G.

; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

; FILE REFERENCE: 21402-416C (Cura-716 SMT)

; CURRENT APPLICATION NUMBER: US/10/210,130

; PRIOR APPLICATION NUMBER: 60/309,501

; PRIOR FILING DATE: 2001-08-02

; PRIOR APPLICATION NUMBER: 60/316,508

; PRIOR FILING DATE: 2001-08-31

; PRIOR APPLICATION NUMBER: 60/354,655

; PRIOR FILING DATE: 2002-02-05

; PRIOR APPLICATION NUMBER: 60/310,291

; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 60/383,887
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/310,951
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/323,936
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: 60/381,039
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 84
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-210-130-84

Query Match 11.6%; Score 994; DB 12; Length 827;

Best Local Similarity 27.3%; Pred.No. 2.5e-73;

Matches 277; Conservative 136; Mismatches 287; Indels 314; Gaps 28;

QY 368 PATPEEGQIIRGWLERSRYGLQAGHNWFIISMOWQWQWKEYVKYDANVPVIEPSSVLN 427
Db 33 PGLDSQWRQIENGESGRER--PLRAGESWFLVEKHWYKOWEAYVQ----- 75
QY 428 GGYKSGTAHPMEQVEDRIGSSLSYVNTTEKFSDNISTASEASETAGSGFLYATPGA 487
Db 76 GG-----DQDSSTF----- 84
QY 488 DVCFAHQHTSDNNNOCLGANGNILLNLPQPGAIIDNOPLVTOBPVKATSLTLEGGL 547
Db 85 -----PCGINNATLFDQ-----INWRL 102
QY 548 KRTPOLIHGRDYEMVPEPVWRALYHWYGANLALPRPVINKNSKTDIP---ELFLFPYLLF 604
Db 103 KE--GLVEGEDYVLLFAAAWHYLVSWYG--LBHGQPIERKVIETLQNKVQVYVPEVLL 158
QY 605 LRQQPATRTQQSNINWVMGNVSPNAPLKRVLAYGCFSEMQTIKEIHEYLSQRLRIK-E 663
Db 159 VRHN-----DLGK-----SHTVQFSHTDSIGLVARTAKERFLVBPQ 194
QY 664 EDMRLWYNSENYLTLLDDEHKLBYLKIQDBQHLYVIEVRNKMDSWPSEMSFIANSKID 723
Db 195 EDTRLWAKNSEGLDLRYDTHITVLDAALETQQLIIMETRKKGDTWPSAQLHVMNNMSE 254
QY 724 RHK-VPTKGTATGLNLTCEMNSSICVSNTPQLTOYFISGRHLYELNRTNPIGMKGH 782
Db 255 EDEDFKGQPGICGLTNLGTCEMNSALQCLSNVPQLTEYFLNLCYLELNFNPLGMKGE 314
QY 783 MAKCYGDLVQELWSGTQKNVAPLKLRTWITAKYAPFNFGQQODSQELLAFLLDGLHEDLN 842
Db 315 IAEAYADLVKQAWSGHRSIVPEHFKNVKGFASQFLGYQHQDSQELLSFLLDGLHEDLN 374
QY 843 RVHEKPYVELKDSGRPDMEVAEAWDNHLRNRSIVVDLPHGQLRSQVKCTCGHISVR 902
Db 375 RYKCKEYVELCDAAGRPDQEAQEAQNEKRENSDIVDTFTHGLFKSTLVCPDCGNVST 434
QY 903 FDPFNFLSLPLPMDSYMHLIEITVILKLD-GTTTPRVYGLRLNMDKVTGLKKQLSDLCGLNS 961
Db 435 PDPFCVLSVPLLIISHKRVLEVFIPMDPRKPEQHRLLVVPKKGKISDLICVALSKTGTISP 494
QY 962 EQILLAEVHGSNIKFPQDNQKVLRSVSGFLCAFEIPVPVSPISASSPTQTDFFSSPSTN 1021
Db 495 ERMMVADVFSR-----FYKLYQLERPLSSIL-----DRD 524
QY 1022 EMFTLTNG-----DLPRPIFIPNGMNTVVPCTEKNKFTNG-----MVNGH--NP 1065
Db 525 DIFVYEVSGRIAEIGSREDIVVPVYLRERTP-----ARDYNNSYGLMLFGHPLLV 576

QY 1066 SLPDSPTCYIIAHRKWMTELYFLSSQKRPSTLFGMPLIVPCTVTRKKDLYDAVWQ 1125
Db 577 SVPRDRFTW-----EGLYNVLMYR 595
QY 1126 VSR-----LASLPPO--BASNAQDC--- 1145
Db 596 LSRVYTKNSDDDDDEKDDKDDVPGSTGSLRDEPEQAQSSGVNRCFPLL 655
QY 1146 DDSMG-YQYP-----FTLVVKDGNCSAWCWYRFCRGCKIDCGEDRA-----FIG 1191
Db 656 DNLGTSQWPPRRRKQLFTLTQVNSNCTS-----DRTTSPVEVHA 696
QY 1192 NAYIAVDWDPALHLRYQTSQERVVDEHESVQSRAQAEPINLDSCLRAFTSEELGEN 1251
Db 697 QPYIALDWEPEMKRYDEVERAEGVYKDCVYWK--KAPVRLQECLELFTVETLEKE 754
QY 1252 EMYYSCKTHCKLAKKLDLRLPILIIHLKRFQFVNGRWIKGKIVKFPRES 1305
Db 755 NPWYCPCKQHOLATKGLDLMLPEILIIHLKRFSTYKFSREKLDLVEFPIRS 808

RESULT 10

US-09-764-881-153
; Sequence 153, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT207
; CURRENT APPLICATION NUMBER: US/09/764,881
; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 153

; LENGTH: 173

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (162)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-881-153

Query Match

Best Local Similarity 10.4%; Score 890; DB 11; Length 173;

Matches 165; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 117 ERMHVVDGKVPDTRLKCFSEGEKYNKFRNWLFLNKDAFTFSRWLLSGVYVLTDD 176
Db 1 ERMHVVDGKVPDTRLKCFSEGEKYNKFRNWLFLNKDAFTFSRWLLSGVYVLTDD 60
QY 177 DPTFFYQTLAGVTHLEESDIIDLEKRYLLKAQSGRTGRFDLETGFLVSPPIRPSLSEGL 236
Db 61 DPTFFYQTLAGVTHLEESDIIDLEKRYLLKAQSGRTGRFDLETGFLVSPPIRPSLSEGL 120
QY 237 FNAFDENRDNHIDFKEISCGLSACCRGFLAERQKFCFKVFDVDRDG 282
Db 121 FNAFDENRDNHIDFKEISCGLSACCRGFLAERQKFCFKVFDVDRDG 166

RESULT 11

US-10-242-747-153

; Sequence 153, Application US/10242747

; Publication No. US20040005577A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT207C1

; CURRENT APPLICATION NUMBER: US/10/242,747

; CURRENT FILING DATE: 2002-09-13

; PRIOR APPLICATION NUMBER: 09/764,881

; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 153
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (162)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

Query Match

Best Local Similarity 10.4%; Score 890; DB 12; Length 173;

Matches 165; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 117 ERMHVVDGKVPDTRLKCFSEGEKYNKFRNWLFLNKDAFTFSRWLLSGVYVLTDD 176
Db 1 ERMHVVDGKVPDTRLKCFSEGEKYNKFRNWLFLNKDAFTFSRWLLSGVYVLTDD 60
QY 177 DPTFFYQTLAGVTHLEESDIIDLEKRYLLKAQSGRTGRFDLETGFLVSPPIRPSLSEGL 236
Db 61 DPTFFYQTLAGVTHLEESDIIDLEKRYLLKAQSGRTGRFDLETGFLVSPPIRPSLSEGL 120
QY 237 FNAFDENRDNHIDFKEISCGLSACCRGFLAERQKFCFKVFDVDRDG 282
Db 121 FNAFDENRDNHIDFKEISCGLSACCRGFLAERQKFCFKVFDVDRDG 166

RESULT 12

US-10-158-057-300

; Sequence 300, Application US/10158057

; Publication No. US20040014039A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: P205C1

; CURRENT APPLICATION NUMBER: US/10/158,057

; CURRENT FILING DATE: 2002-06-12

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 384

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 300

; LENGTH: 173

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (162)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-10-158-057-300

Query Match

Best Local Similarity 10.4%; Score 890; DB 12; Length 173;

Best Local Similarity 99.4%; Pred. No. 7.5e-66; Matches 165; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 117 ERMHVVDGKVPDILRKCFSEGEKYNVEKFNWFLFNKDAFTFRWLLSGGVVYTLTDDS 176

Db 1 ERMHVVDGKVPDILRKCFSEGEKYNVEKFNWFLFNKDAFTFRWLLSGGVVYTLTDDS 60

QY 177 DTPFYQTLAGVTHLESDIIDLEKRYWLLKAQSGRTGREDLETFGPLVSPPIRPSLSEGL 236

Db 61 DTPFYQTLAGVTHLESDIIDLEKRYWLLKAQSGRTGREDLETFGPLVSPPIRPSLSEGL 120

QY 237 FNAFDENRDHIDFKEISCGLSACCRGFLAERQKFCFKVFDVDRDG 282

Db 121 FNAFDENRDHIDFKEISCGLSACCRGFLAERQKFCFKVFDVDRDG 166

RESULT 13

US-09-764-875-1064

; Sequence 1064, Application US/09764875

; Publication No. US20040018969A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: P202

; CURRENT APPLICATION NUMBER: US/09/764,875

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1249

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1064

; LENGTH: 173

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (162)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-875-1064

Query Match

Best Local Similarity 10.4%; Score 890; DB 12; Length 173;

Matches 165; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 117 ERMHVVDGKVPDILRKCFSEGEKYNVEKFNWFLFNKDAFTFRWLLSGGVVYTLTDDS 176

Db 1 ERMHVVDGKVPDILRKCFSEGEKYNVEKFNWFLFNKDAFTFRWLLSGGVVYTLTDDS 60

QY 177 DTPFYQTLAGVTHLESDIIDLEKRYWLLKAQSGRTGREDLETFGPLVSPPIRPSLSEGL 236

Db 61 DTPFYQTLAGVTHLESDIIDLEKRYWLLKAQSGRTGREDLETFGPLVSPPIRPSLSEGL 120

QY 237 FNAFDENRDHIDFKEISCGLSACCRGFLAERQKFCFKVFDVDRDG 282

Db 121 FNAFDENRDHIDFKEISCGLSACCRGFLAERQKFCFKVFDVDRDG 166

RESULT 14

US-10-369-493-22797

; Sequence 22797, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 22797

; LENGTH: 979

; TYPE: PRT

; ORGANISM: Schizosaccharomyces pombe

US-10-369-493-22797

Query Match

Best Local Similarity 9.2%; Score 787.5; DB 12; Length 979;

Matches 283; Conservative 162; Mismatches 414; Indels 377; Gaps 37;

QY 432 SFGTAAPMEQVEDRIGSLSVNTTEKFSNISTASEASTAGSGLYSATPGADVCF 491

Db 19 SLSESSQSSSNVDDISQKSIISLGDAE--ISKVLSIAEQKQLIG----- 61

QY 492 ARQHTSDNNOCILGANGNILLHNPQK-----PGADNPFLVTOEPVKATS 539

Db 62 ----ELVNNQPELELQGVNDYILSYWYERLCYLAEDGPFPGVDQEDIA----- 108

QY 540 LTLGGRLKRTPLIHGRDYEMVPEPVWRYALVHWYG--ANLALPRPVIKNSKTDIPEL--E 596

Db 109 -DLGTGLK--PDLOEIDFTIISRDVMDLLVRYWYGLKGPFPFRETVMNGSSEHPLVVE 165

QY 597 LFPR-YLLFRLQCPATRTQQSNINWYMGVSPNAPLKRVLAYTCFSRMQTIKEIHYL 655

Db 166 VYPIPIFSLTLLSTNAVDANES-----HKPKKISL-----SSKSTLEDLEGV 207

QY 656 SQRLAKSEDAFLWYNSENVLTLLDDHDKLEYLKIQDBQHLVIEVRNKM----- 707

Db 208 KYTSLSPDQFELRVDTDQPLHRTIDPS---SFIKINSKEIIDFLEKSKTLVELGMDSS 264

QY 708 -----SWPEE-----MSFI-----ANSKIDRHKVPTKGGATGLSNIGNTCF 744

Db 265 CSLVAECMNETWVDRLALRQLQFLIQORNNSSNEEQKEKVP---GTCGLSNIGNTCY 321

QY 745 MNSSTQCVSNTQPLTQYFISGRHLYELARTPIGMKGMKACYDGLVQELMSGTKQNVAP 804

Db 322 MNSALQCLTHTRELDRDFTSDEWKQVNESNPLMGVGQGVAFASLIKSLYSPEHSEAP 381

QY 805 LKLRWTIAKYAPRFGFQQSQOELLAFLLDGLHEDLNRVHEKPVV---ELKDSGRPDW 861

Db 382 RQKATIGKFNHISFLGYCQQSQOEFLLFLLDGLHEDLNRIYQKPYTSKPDLEVEDEEK 441

QY 862 EVAABAWDNHLERNKSIIVDLFHGQLRSQVCKTCGHISVRFDPFNFLSLPLMDSMYHL 921

Db 442 NTAECWRLHLRNDLSLIVDLFGQMYRSTLVCPCVNTVSIITFDPMDLTLPVVKVWNSH 501

QY 922 EITVTKLD-GTTPVRYGLRLNDE-----KYTGKQLSDLCGLNSEQILLAEVHGSN 973

Db 502 TVTFIPADTNLTPLAIVLVLESKAATIEDLVKYVAEKSGCSDY-----RKILVTETYKGR 556

QY 974 IKNF-PQDNQKVRLSVSG---FLCAFEIP-----VPVSPISASSPTQD--FSSSP 1018

Db 557 FYRFLTQLSKSLMEISEDEIYLYELRYPYEDGSDDDILVPVYHISDDSTNSANSYMSR 616

QY 1019 STNEMFTL-TTNGDLPRFIFIPNGMNTVPCGTEKNFTN-----GMVNGHMPSLPDSPP 1072

Db 617 DFGHPFLQLSDNEVTDASFISEKLKLYQQOFTLLKLNKIDSLESLELGH-----EDQV 672

QY 1073 TGYIIAVHRKQWRTLEYFLSSQKQR---PSLFGWPLI-VPCTVHTRKKOLYDAWVQVS 1127

Db 673 QKGPLDVMDHDSQTFLEMRVHDFRFEKIPFGWNMSVSNLFLTERDKO----- 722

QY 1128 RLASPLPPOEASNAHQDCDDSGYQYPTFLRVQKDGNSCAMCPWYRFRGCKIDCGE-D 1186

Db 723 -LESTVDPDLDAHSIEBEDSEFKDVAPOGYPSPKSNEN-----TKLTAKEND 769

QY 1187 RAFIGNAVIADWDPTALHLYQT-----SQERVDEHESVQSRRAQAE 1232

Db 770 RLIIQGDLLVCEWPEKSYQFVFSVAPSPQWGRSLWLESKTLSDKDDSDSR----- 824

QY 1233 INLDSCLRAFTSEEBELGENEMYCCKTHCLATYKLDLMLPILLIHLKRFQFVNGRW 1292

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:36:36 ; Search time 28 seconds

(without alignments)

5509.092 Million cell updates/sec

Title: US-09-888-615-73

Perfect score: 8563

Sequence: 1 MGAKERIGFLSYEALRRV.....DTSSMDDEDFDYKXYCVLQ 1604

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Pir1.*
- 2: Pir2.*
- 3: Pir3.*
- 4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4531	52.9	1089	2	S22158 transforming prote
2	1406	16.4	786	2	S22155 oncogene 1 (tre-2
3	1124	13.1	1041	2	T29010 hypothetical prote
4	1087	12.7	963	2	T09478 ubiquitin thiolest
5	1006	11.7	937	2	T04194 hypothetical prote
6	981.5	11.5	887	2	C86453 CDS protein F9L11.
7	980.5	11.5	794	2	I58376 hypothetical prote
8	967.5	11.3	928	2	T04192 hypothetical prote
9	920.5	10.7	914	2	T00757 probable ubiquitin
10	787.5	9.2	909	2	T41006 ubiquitin carboxyl
11	756.5	8.8	900	2	T33734 hypothetical prote
12	738.5	8.6	1254	2	S46636 hypothetical prote
13	693.5	8.1	849	2	T41085 probable ubiquitin
14	675	7.9	1095	2	T20528 hypothetical prote
15	371	4.3	926	2	S39344 deubiquinating enz
16	361.5	4.2	1083	2	T00790 ubiquitin-specific
17	330	3.9	805	2	S50277 ubiquitin-specific
18	330	3.9	961	2	H86181 hypothetical prote
19	313.5	3.7	268	2	T46902 hypothetical prote
20	303	3.5	438	2	T39772 hypothetical prote
21	259	3.0	445	2	I46225 ubiquitin carboxyl
22	249.5	2.9	1071	2	S48378 probable membrane
23	239	2.8	557	2	T50798 ubiquitin specific
24	211.5	2.5	717	2	S38177 SSV7 protein homol
25	203	2.4	457	2	T41069 ubiquitin carboxyl
26	203	2.4	555	2	T23531 hypothetical prote
27	200	2.3	526	2	JC6133 deubiquitinating e
28	190	2.2	754	2	S50601 hypothetical prote
29	184.5	2.2	875	2	T50182 ubiquitin-specific

ALIGNMENTS

RESULT 1

S22158

transforming protein (clones 210 and 213) [imported] - human

N.Alternate names: oncogene 2 (tre-2 locus)

C.Species: Homo sapiens (man)

C.Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 03-Nov-2000

C.Accession: S57868; S57875; S22156; S22158

R.Nakamura, T.; Hillova, J.; Mariage-Samson, R.; Onno, M.; Huebner, K.; Cannizzaro, L./ Oncogene 7, 733-741, 1992

A.Title: A novel transcriptional unit of the tre oncogene widely expressed in human car

A.Reference number: S57867; MUID:92228503; PMID:1565468

A.Accession: S57868

A.Status: preliminary

A.Molecule type: mRNA

A.Residues: 1-1089 <NAK1>

A.Cross-references: EMBL:X63547; NID:G37332; PIDN:CAA45111.1; PID:G37334

A.Experimental source: Clone 213

A.Accession: S57875

A.Status: preliminary

A.Molecule type: mRNA

A.Residues: 504-1089 <NAK2>

A.Cross-references: EMBL:X63546; NID:G37329; PIDN:CAA45109.1; PID:G37331

A.Experimental source: Clone 210

A.Note: submitted to the EMBL Data Library, December 1991

Query Match 52.9%; Score 4531; DB 2; Length 1089;
Best Local Similarity 93.7%; Pred. No. 9.2e-284;
Matches 851; Conservative 22; Mismatches 31; Indels 4; Gaps 1;

QY	701	EVRNKMDSWP	EMSP	STANSS	KIDRHKVTEK	ATGLSLN	GNLNT	CFMNSS	IQCVSN	TOPLTQ	760
DB	182	EVNKMDSWP	EMSP	STANSS	KIDRHKVTEK	ATGLSLN	GNLNT	CFMNSS	IQCVSN	TOPLTQ	241
QY	761	YFISGRHLYEL	NRNTP	ICMGKMAK	YCDLV	QELWSG	TQKNV	APLKL	EWTTIAK	YAPRNG	820
DB	242	YFISGRHLYEL	NRNTP	ICMGKMAK	YCDLV	QELWSG	TQKNV	APLKL	EWTTIAK	YAPRNG	301
QY	821	FOQDSQEL	LAFLDGL	HEDLNR	VHEKPYVEL	KDSGD	GPDMWE	VAEAWD	NHLRRNS	IVV	880
DB	302	FOQDSQEL	LAFLDGL	HEDLNR	VHEKPYVEL	KDSGD	GPDMWE	VAEAWD	NHLRRNS	IVV	361
QY	881	DLFHGQLASQ	VKCKTC	GHSIVR	FDPPNF	LSLP	MDSYMH	LEITV	IKDGTTP	VRYGLRL	940
DB	362	DLFHGQLASQ	VKCKTC	GHSIVR	FDPPNF	LSLP	MDSYMH	LEITV	IKDGTTP	VRYGLRL	421
QY	941	NMDEKYG	TKGKQL	SDLCGL	NSEQILL	AEVHGS	NTKN	FPQDNQ	KVRLSV	SGFLCAFEI	1000
DB	422	NMDEKYG	TKGKQL	SDLCGL	NSEQILL	AEVHGS	NTKN	FPQDNQ	KVRLSV	SGFLCAFEI	481
QY	1001	VSPISAS	PTQIDF	SSSPST	NEMFT	LTNGDL	PRPI	FIPNGM	PNVTVP	CGCKNF	1060
DB	482	SSPISAS	PTQIDF	SSSPST	NEMFT	LTNGDL	PRPI	FIPNGM	PNVTVP	CGCKNF	541

probable ubiquitin
hypothetical prote
frequentin - Africa
neurocalcin - fruit
related to neurona
hypothetical prote
protein T05H10.1
hypothetical prote
guanylate cyclase-
hypothetical prote
probable ubiquitin
neurocalcin (clone
guanylate cyclase
hypothetical prote
ubiquitin-specific
hypothetical prote

QY	1061	NGHMPSLDPSPTGYIIAYHRKMRTELFLSSQKNRPSLFGNPLIVPCTVTRKKDLVD	1120
Ddb	542	NGHMPSLDPSPTGYIIAYHRKMRTELFLSPQENRPSLFGNPLIVPCTVTRKKDLVD	601
QY	1121	AWIIOVSRLASPLPOEASNHAQDCDDSGYGYPTFLRVVQKGNSCAWCPWRRCGCK	1180
Ddb	602	AWIIOVSWLARLPPOEASHAQDRNCNGYQYPTFLRVVQKGNSCAWCPYRCRGCK	661
QY	1181	IDCGEDRAPIGNAYIAVDMDPTALHLRYOTSQBRVVDESHESVEQSRRQAQAEINLDSCLR	1240
Ddb	662	IDCGEDRAFIGNAYIAVDMDPTALHLRYOTSQBRVVDKHESVESRRAQAEINLDSCLR	721
QY	1241	AFTSBEELGENEMYCSKKTCHCLATKGLDLWRLPPIILHLKRFOYNGRWIKSQKIVK	1300
Ddb	722	AFTSBEELGESEMYCSKKTCHCLATKGLDLWRLPPIILHLKRFOYNDOWIKSQKIYR	781
QY	1301	FPRESFDPFAFLVPRPALCQHKPLTPQGDLSLSEPRILAREVKVQDAQSSAGEEDVLISK	1360
Ddb	782	FLESFDPFAFLVPRPALCQHKPLTPQGDLSLSEPRILAREVKVQDAQSSAGKEDWLISK	841
QY	1361	SPSSLGANITISPPKGPSSRSKGTSCPSKSNSSPNSSPRTIGRSKRURLPQIGSKNKL	1420
Ddb	842	SPSSLGANITSSPPKGPSSRSKGTSCPSKSNSSPNSSPRTIGRSKRURLPQIGSKNKP	901
QY	1421	SSSKENLDASKENGACQICELADALSRGHVLGSGSQPELVTPQDHEVALANGFLYEHCAG	1480
Ddb	902	SSSKENLDASKENGACQICELADALSRGHVRGGSQPELVTPQDHEVALANGFLYEHCAG	961
QY	1481	N-----GYSQGQLGNHSESDTDQREDTRIKPIYNLYAISHSGILGGHYITYAKNPNC	1536
Ddb	962	NGCCDGYSQGQLGNHSESDTDQREDTTHIKPIYNLYAISHSGILGGHYITYAKNPNC	1021
QY	1537	KWCYNDSSCKELHPDEIDTDSAYILFYQQGDIYAQFLPKTDGKMADTSMDDEDSED	1596
Ddb	1022	KWCYNDSSCEELHPDEIDTDSAYILFYQQGDIYAQFLPKIDGKMADTSDEDESSED	1081
QY	1597	YKXYCVLQ 1604	
Ddb	1082	YKXYSMLQ 1089	
 RESULT 2 S22155			
oncogene 1 (tre-2 locus) (clone 210) - human			
N;Alternate names: transforming protein			
C;Species: Homo sapiens (man)			
C;Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 05-Nov-1999			
C;Accession: S57874; S22155			
R;Nakamura, T.; Hillova, J.; Mariage-Samson, R.; Onno, M.; Huebner, K.; Cannizzaro, L.A.			
Oncogene 7, 733-741, 1992			
A;Title: A novel transcriptional unit of the tre oncogene widely expressed in human cancer			
A;Reference number: S57867; MUID:92228503; PMID:1565468			
A;Accession: S57874			
A;Status: preliminary			
A;Molecule type: mRNA			
A;Residues: 1-786 <NA>			
A;Cross-references: EMBL:XG3546; NID:g37329; PIDN:CAA45108.1; PID:g37330			
 Query Match 16.4%; Score 1406; DB 2; Length 786;			
Best Local Similarity 92.7%; Pred. No. 2.2e-82;			
Matches 266; Conservative 7; Mismatches 14; Indels 0; Gaps 0;			
QY	701	EVNKOMSWPEEMSFANSKIDRHKVTEPKGATGLSNLONTCFMNSSIOCVSNTQFLTQ	760
Ddb	499	EVNKOMSWPEEMSFANSKIDRKQVTEPKGATGLSNLONTCFMNSSIOCVSNTQFLTQ	558
QY	761	YFTSGRLHYELNRNPIGMKGHWAKCYGDLVOELWSCTOKNVAPLKLRWTIAKYAPRENG	820
Ddb	559	YFTSGRLHYELNRNPIGMKGHWAKCYGDLVOELWSCTOKSVAPLKLRWTIAKYAFPDG	618
QY	921	FOQODSQELLAFLLDGLDHEDLNRYEHXPYYVELKDSDGRPDWEVAEAZWNLRRNSIVV	880

Db 410 WQVVKNEBGNVLDDGAQNLHOLYSLSGLTKKKVKNKMLLEVRGRTGVWPEELRASUS 469
QY 714 --SFIANSKXIDRHKVPTEKATGLNLGNTCFMNSSIQCVSNTOPLTOYFISGRHLYEL 771
Db 470 GKQITAASTLSSNAQLSGRPGAVLNGYGNFYRNAGIQCLARVSPLTOYFLDEDNLDAI 529
QY 772 NRTNPIGMKGMKAKC---YGDIVQELWSGTQKQVAPLKLRTWTIAKYPFRNGFQQQDSOE 828
Db 530 KRGNL--RGDAAEETIEVAKLREXWAAKQNIAP----- 563
QY 829 LLAFLLDGLHEDLNVRHEKPYVELKDS--GRPDWEVAAEAWDNHLRNRISIVVDLPHGOL 887
Db 564 -----NDL-----IKKKEADEADEEKAERSWTEYEKQNESLTVQFTGOL 605
QY 888 RSQVKCTCGHISVRDPNFIPLPMDISYMHLEITVIKLDGTPVRYGLRLNMDKTY 947
Db 606 RSKLIRCTQSSSVPEPTSLSPICFEDVDLYQIVVHRDGRIPRRYGFLSRDSKVG 665
QY 948 GLKKQLSDLCGLNSEQIILAEVHGSNIKFPQDNQKVLRSVSGFLCAFEIPVPVSPISAS 1007
Db 566 NUREVAVSGISMSHLTQ-----CMSSKGTL-----MSRSPNHS 702
QY 1008 SPTQTD--FSSSPSTNEMFTLTNGDLPRPIFIPNGMPTVVPCTGKFTNGVMNGHMP 1065
Db 703 SNLRDELPLSSPFGARLYAL----- 723
QY 1066 SLPDSPTFG--YIIAVHRKM--RTELYFLSSQKN--RPSLFGMPLIVPCTVHTRKKDLYD 1120
Db 724 ELPES--TSDQWRVAMHRKQJNHPEYILGSTAGFIVSRFGLPLIVGLDEEVTOKLYE 781
QY 1121 AVWQVSRLASPLPPOEASNAHQDQDSDSGYQYPTLVRVQDGNSCAWCPWYRCRCK 1180
Db 782 DVMYQHRFWHSVNSSSRAHDPCDE--NSGYPTLCVLDVNYEWCGCPALRCRCP 840
QY 1181 IDCGEDRAFI--GNAYTAVDWDPTALHLRVTQSERVVDHEHSVEOSRAQAPINLDSCL 1239
Db 841 IRPDESKVIPANCPTAVDMLPFIALYLYNHSQEQACEDDPSVAETGRHFPAPSSLEHCI 900
QY 1240 RAFTSEEEGENEMYCSKCKTHCLATKLDLWRPLPILIHLKRFQV--NGRWIKSQK 1297
Db 901 EKFSCPETL--DAAIQCDCEKKTWRDKWMTWKLPKYLIIHLKFEFLRQGRMGKCR 958
QY 1298 IVKFPRESDDPAFLVRDPALCOHKLPTQGDDELSEPRILAREYKVKYDAOSSAGEEDVL 1357
Db 959 TVNFFLKHDPAPP----- 972
QY 1358 LSKSPSSLANIIISPKGSPSSRSKSGTSCPSKSNSSPNSPRTLGRSKGRILRPLQIGSK 1417
Db 973 -----VDKPDGN----- 979
QY 1418 NKLSSKENLDASKENGAGQICELADALSRGHVGGSOPELVTPQDHEVALANGFLYEHE 1477
Db 980 ----- 979
QY 1478 ACGNGYNGQLGNHSEEDSTDQREDTRIKPIYNIYALISCHSGILGGHYTYAKNPCK 1537
Db 980 -----TYECIALANHYQLSCGHHFIAYAKSNEDK 1008
QY 1538 WYCYNDSSCKELHPDEIDTDSAYILFYEQGI 1569
Db 1009 WLLNDCSVRESEEBVDKQAYLLFYERKDV 1040

RESULT 4

T09478
N:ubiquitin thiolesterase (EC 3.1.2.15) - human
C:Species: Homo sapiens (man)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T09478
R:Gray, D.A.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z16684

A:Accession: T09478

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-963 <GRA>

A:Cross-references: EMBL:U20657; NID:g2459396; PID:g2459395

C:Genetics:

A:Gene: Unph

A:Map position: 3

C:Keywords: proto-oncogene; thiolester hydrolase

Query Match 12.7%; Score 1087; DB 2; Length 963;

Best Local Similarity 23.9%; Pred. No. 1.1e-61;

Matches 327; Conservative 168; Mismatches 313; Indels 560; Gaps 40;

QY 367 RPATPEEGQTIIRGLERESRYGLQAGHWNFIISMOWQWQKVKYDANPVVIEPSSVL 426

Db 10 RPDAAETQKSEL--GPLMRTT--LQGAQWYLIDSRWFKQWKYVGF----- 52

QY 427 NGGKYSFGTAHPMEQVEDRIGSSLSYVNTTERKFSNISTASEASETAGSGFLYSATPG 486

Db 53 SWDMYNYG----- 60

QY 487 ADVCFARQHNISDNNNQCLLGANGNILLNLPQKGAIDNOLVTOEPVKATSLITLEGGR 546

Db 61 -----EHLNF-----PGIDNSGLFS--DP-----ESQT 82

QY 547 LKRTPOLIHRGDYEMVPEPVWRALYHYWGANTALPRPVIKNSKTDIPELELFP----- 600

Db 83 LKE--HLIDELDYLVPTAANKLLNMYGC--VEGQQPIVRK-----VVEHGLFVKHCKVEV 135

QY 601 YLLFLRQCPATRTQQSNINWYMGVSPNAPLAKRLAYTGC--FSRMTIKIEIHEYLSQRL 659

Db 136 YLLEKL-----CENSDDPTNVL-----CHFSKADTIATIEKMRKLF 173

QY 660 RI--KEEDMRLWLYNSENLYLTLDDDEHKLEYLKIODEHLVIEVRNKMSPWEE----- 712

Db 174 NIPAEERETLWKNYMSNTVEQLSKLNTVQDAGLYQGVLVIEPQNEGTGTPQTLOSRS 233

QY 713 -----MSIAN-----SSKIDR-----HKVP 728

Db 234 STAPSNFTTSPKSSASPYSSVASLIANGDSTSTCGMHSSGVSRGSGSFSYNOQEP 293

QY 729 T--EKGATGLSLNGTCFMNSSIQCVSNTOPLTOYFISGRHLYELNRTNPIGMKGMKAK 785

Db 294 SSIHQFGLCGLNLGNTCFMNSALQCLSNLTAPLDYFLKDEYEAENRDNPLGMKEIAE 353

QY 786 CYGDLVQELWSGTQRNVAPLKRWTIAKYPFRNGFQQQDSQELLAFLDGLHEDLNVRH 845

Db 354 AYAEELIKOMWSGERDAHVAPSMFTQVGRFAPQPSGYQQQDSQELLAFLDGLHEDLNVRK 413

QY 846 EKPVELKDSGRPDWEVAAEAWDNHLRNRISIVVDLPHGQLRSQVKCTCGHISVRPDP 905

Db 414 KKPYLELKXANGRPDAVAKAEWENHRLRNDSVIVDTFHGLFKSLVCPKAVSVTFDP 473

QY 906 FNFSLPLPMDSYMHLEITVIKLD--GTPVRYGLRLNMDKTYTGLKKQLSDLCGLNSEQI 964

Db 474 FCYLTPLPLKKORVMEVELVPADPHCRPTQYRTVPLMGAVSDLCCEALSRLSGIAAENM 533

QY 965 LLAEVHGSNI--KNFPQDNQKVLRSVSGFLCAFEIPVPVSPISASSPTQTDFFSSSPSTNEM 1023

Db 534 VVADVNRHFRHKIFQMDGLNHPRDDIFVYEV-----CSTSVDSGSEC 577

QY 1024 FTLTNGDLPRPIFIPNGMPTVVPCTGKFTNGVMNGHMPSLPDSPTGTGTIAVHRKM 1083

Db 578 VTL-----PV----- 582

QY 1084 MRTELYFLSSQKNRPS-----LFGMPLIVPCTVH--TRKKDLYDAVWTOVSRRL----- 1129

Db 583 -----YFERKSRPSSTSSALYQPLLSVPKHLTLESLYQAVCDRIERYVKQPLP 636

QY 1130 ---ASPLPPEQASNAHQDC---DDSMGYQYPTLVRVQDGNSCAWCPWYRCRCK 1181

Db 637 DEFGSSFLPFGACNRSNCSCEGEDEEHEHQEGKQLSETEGS----- 680

```
QY 1182 DCGED-----RAF-----IGNAY-----I 1195
DB 681 --GEDEPNDSPTQKKIGQPCPKRLFTPSLVNSYGTADINSLADGKLLKLNRSSTL 738
QY 1196 AVDMDPTALHLRYQTSQBRVDEHESVQSRRAQAEPINLDSCLRAFTSEBELGENEMY 1255
DB 739 AMDWDRETRRLYYDQSEAEVKEHVSMLQPKKKKTTVALRDCIELFTTMTLGEHDPWY 798
QY 1256 CSKCTHCLATKCLDLWRLPPIILHLKRFQVNGRW-IKSQKIVKPPRESFDSAFLVP 1314
DB 799 CNCKKHOATKFDLWLSPLKLVVHLARFSY-NRYWRDKLDTVVEPPIRGLNMSF--- 854
QY 1315 RDPALCQHKPLTPQGDDELSEPRILAREVKVDAQSSAGEEDVLLSKSPSLSANIISPK 1374
DB 855 ----- 854
QY 1375 GSPSSSRKSGTSCPSKSNSSPNSSPRTLGRSKGRRLRPLQIGSKNKLSSKENLDAKENG 1434
DB 855 ----- 854
QY 1435 AGQICELADALSRGHVGLGSGQPELVTPODHEVALANGFLYEACGNGYSGQLGNHSEE 1494
DB 855 ---VCNLS----- 859
QY 1495 DSTDDQREDTRIP-IYNLYAISCHSGILGGHYVYTAKNP-NCKWYCYNDSSCKELHPD 1552
DB 860 -----ARPYVYDLIAVNSHYGANGVGHYTAFAKNGKNGWYVYFDDNSVLASED 908
QY 1553 EIDTDSAXILFYEQGIDYAOPLKTDGKQVAD---TSSWDEDPESD 1596
DB 909 QIVTKAAVYLFYORDDDEFYK-TPSLSSSGSGSDGGRFPSSQGGFGDD 955

RESULT 5
T04194
hypothetical protein T4F9.50 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
C:Accession: T04194
R:Bevan, M.; Wedler, H.; Wedler, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15260
A:Accession: T04194
A:Molecule type: DNA
A:Residues: 1-937 <BEV>
A:Cross-references: EMBL:AL049523
A:Experimental source: cultivar Columbia; BAC clone T4F9
C:Genetics:
A:Map position: 4
A:Introns: 51/2; 122/2; 170/3; 190/3; 250/3; 375/3; 422/3; 479/3; 565/3; 617/2; 795/2; 8
A:Note: T4F9.50

Query Match 11.7%; Score 1006; DB 2; Length 937;
Best Local Similarity 24.8%; Pred. No. 1.8e-56;
Matches 321; Conservative 153; Mismatches 336; Indels 482; Gaps 38;

QY 368 PATPEEGOIRGLWERSRYGLOAHNWFILSMQWQWKEYVYKDYANPVVIFSSVLN 427
DB 18 PTPPEEKIV-SELITEEDNLKGNLYFVISKWYTSWEKIVE----- 61
QY 428 GGYSGFTAHPMEQVEDRIGSSLSVNTTEKFSNDISTASEASETAGSGFLYSATPGA 487
DB 62 -----QSTKEYISGESSEAS----- 76
QY 488 DVCFAQRTNSDNNQCLLGANGNILLHNLPKQPAINDQPLVTOEPVKASLTLEGRLL 547
DB 77 -----RPGFIDNHDISES-----DVNDPQL 98
QY 548 KRTPOLIGHRDYEMVPEPVYKALYHWYGANLALPRPVT-KNSKTDIPLELFPYLLFLR 606
DB 99 RRL--LMERVDIVLPQEWKRLVENVYSGGPIERKLCQGFYTSYVSVYFICLMLTD 156
```

```
QY 607 QQPATFTQCSNIWANGVNPSPNAPLKRVLAVTGCFSRMQTIKEIHEVLSORLRKEBDM 666
DB 157 GRDESRTV-----IRLGKQASIRELYEKVCAUTG-----VPQKFLMKDALRYEDF 204
QY 667 -----RLWLYNSENYLLTLDDEPHK-LEVLIKIOEQLHVLIEVRNKMOMPE 711
DB 205 AALPHIDIPFYQAIHWDFDKRKNGLDLSYSKSLERSSLHMDQDILLEVDSGSSS-QS 263
QY 712 EMSFIANSKI-----DRHKV-----PT----- 729
DB 264 AMSSTGNEALVPLEPSRSSVTIAGPTLSNGHSTSNPSLPPRITSEDDGNSLSILCK 323
QY 730 -EKGA-TGLSNLGNCTCFMNSSIQCVSNTPQLPYFISGRHLYELNRTNPIGMKGMWAKY 787
DB 324 GEKGLAGLNLGNTCFMNSALQCLAHTPIVEYFLQD-YSDDINRDNPLGMCGLAIATF 382
QY 788 GDVLQELWSGTQKNVAPLKLWRTIAKAPRFRNGFOQSOELLAFLLDGLHEDLNVRHEK 847
DB 383 GDLLKLLWSGSRNSVAPRAFKTKLARFAPQFGYQNHDSQELLAFLLDGLHEDLNVRK 442
QY 848 PYVELKDSGRPDWEVAAEAWNDNHLRRNSIVVDLFGQLRQSRQVCKTCGHSISVRPDPFN 907
DB 443 PYIELKDSRDPDEVAEELWNYHKAENDSVIVDCQGYKSTLVCPACGKISITFDPDM 502
QY 908 FLSPLPMDSYMHELTIVKLDGT-TPVRYGLRLNMDKEYTGLKQLSDLQGL-NSEQIL 965
DB 503 YLSVPLSTLTSMVTVFYCDGSLPMEYTVIVFKNGSIRDILITALTGTACLLAEDESLL 562
QY 966 LAEVHGSNKNPQDNQKVLRSVSGFLCAFEIPVPVSPISASSPTQTPSSSPSTNEMPT 1025
DB 563 LAEVYDHKIFKY-----FENPL-----DLSLSIKDDEHIV 592
QY 1026 LITNGDLPRPIFIPNGMNTVVPQGTERTNGMNGHMPSLPDSFFTGYIITAVHKKMR 1085
DB 593 AVELNQMPK---GSGKAKLEILHGGQKRPILSVRGRDVKLFGTPFVTYV----- 639
QY 1086 TELYFLSSQKNRPSLFGMLPIVPCVHTKKDLYDAVMIQVSRKLSPL---PPOEASNH 1141
DB 640 -----NTEPLSGADI-----DAV---LSRFLSPLKHVAPSKIING 672
QY 1142 AQ-----DCDDSMGYQPTFLRVVQDGNSCAWCPWYRRCGCK 1180
DB 673 SENGHLPDATVDEASEILSSPDTEDDASRELSP-RIFLTDERGLNFKPLQ----- 723
QY 1181 IDCGEDRAFIGNA-YIADWDPTALHLRYQTQSERVWDE-HESVEQSRRAQAEPINLDS 1238
DB 724 ---SESSI SLGTATRYLVLEWN-EGEHRYDSSYLSOLPEVHKTSFSAKKTROESISLFS 779
QY 1239 LRAFTSEELGENEMYCYCKCKTHCLATKKLDLWRLPPIILHLKRFQVNGRWIKSQ-- 1296
DB 780 LEAFLAEELPGDDMWFCPSCKEHRQANKKLDLWKLPIILVFLHAKFTY--SRYLKNKID 837
QY 1297 KIVKFPRESFDSAFLVPRDPAQCQHKPLTPQGDDELSEPRILAREVKVDAQSSAGEEDV 1356
DB 838 TFVNFVPHLDLSKYV----- 853
QY 1357 LLSKSPSLSANIISSPKGPSSSRKSGTSCPSKSNSSPNSSPRTLGRSKGRLLPQIGS 1416
DB 854 ----- 853
QY 1417 KXKLSSENLDASKENGAGQICELADALSRGHVGLGSGQPELVTPODHEVALANGFLYEH 1476
DB 854 KXNKDQS-----YLYE- 864
QY 1477 EACGNGYSGQLGNHSEEDSTDQREDTRIKPIYNLYAISCHSGILGGHYVYTA--NP 1534
DB 865 -----LYAVSNHYGGLGGHYTAYAKLIDD 889
QY 1535 NCKWYCYNDSSCKELHPDSDTDSAYILFYEQ 1566
DB 890 N-EMWHFDDSHVSSVNESEIKNSAAVLYFYRR 920
```

RESULT 6

C86453
 CDS protein F911.5 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C86453
 R:Neologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ket, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: C86453
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-887 <STO>
 A:Cross-references: GB:AE005172; NID:G6910582; PIDN:AAF31287.1; GSPDB:GNC00141
 C:Genetics:
 A:Gene: F911.5
 A:Map position: 1

Query Match 11.5%; Score 981.5; DB 2; Length 887;
 Best Local Similarity 23.5%; Pred. No. 6.4e-55;
 Matches 311; Conservative 153; Mismatches 304; Indels 557; Gaps 38;
 370 TPEEQIIRGLWERSRYGLOAGHNFIIISQWQWQKEYKYDANPVVIBPSSVLNGG 429
 18 TPEERIVTE-LNNEABADLKGNDYFVSNRYWRQRFVGL----- 60
 430 KYSFGTAHPMEQVEDRIGSSLYNNTTEKFSDNISTASEASTAGSGFLYSATPGADV 489
 61 -----LTEE-----FRSGEPSEVT----- 74
 490 CFARQHTSDNNQCLLGANGNILLHLNPKQKCAINDQPLVQEPVKATSLTEGGLR 549
 75 -----RPGIDNDIIDSSEAS----- 92
 550 TPQ-----LIHGRDYMPVPVRALVHYMGANLALPRVKN---SKTDIPELEFPRLV 602
 93 DPQLRWLEEGVDYTLVQEVWKLKVKYKGGPPVKRLISQGFYTKSPSEVYLLCTL 152
 603 LFLRQPARTRQOSNIWVWNGVPSNAPLKRVLAVTGCFSRMOTIKEIHEYLSORLRIK 662
 153 TDSRDESTIIR-----LSQASTGQLYEMV-----CA 180
 663 EEDMLWLYNSNYLTLD-DEDHKLEVLKIOEQHLVIEVRNKMSPMEANSFIANSK 721
 181 GKARLWDYFEKKSVLLDPSSEQSVBEAGLOFNQDILLEVDGSASS-----QFVMSLAE 235
 722 IDRHKVPTE-----RGA 733
 236 NELAMVLEPMRSDAMDIVRGGTLLNGSHNGKPFQFCNRTFKDDVSSRTFGKKEKRL 295
 734 TGLSNLGNTECFWSSIQYNTQPLTOYFISGRHLYELNRTPIGMKGMKACVGLVQE 793
 296 GGLQNLGNTECFWSSIQYNTQPLTOYFISGRHLYELNRTPIGMKGMKACVGLVQE 793
 794 LMSGTOKNAPLKLRTIATYAPRFGFQQQDSQELLAPLGLDGLHEDLNVRHVEKYVELK 853
 355 LWSGGQNTVAPRAFKTLARFAPQFSGYNQHSQEWLAPLGLDGLHEDLNVRHVEKYELK 414
 854 DSDGRDWEABAWNHLRRNSIVVDLPHGLRSQVKCKTCGHSVRFPDFNPLSLPL 913
 415 DSDGRDDEVAEEKWKYKHARNDSDIVDFQGYKSTLVCPCDKISITFDFFMVLSP 474
 914 PMDSYMHLEITVLKDG-TPVRYGLRLNMDEKTYGLKKQLSDLCGL-NSOQILLAEVHG 971
 475 PSSRTSMITVYFGDGHLPMPYTYTVFVKGSCRDLSNALGTACCLDNDESLLLA 534

QY 972 SNIKNFPQDNQKVLVSGLCAFEIPVPVSPISASSPTQTDSSSPSTNEMFTLTNGD 1031
 DB 535 HKVFKYEN----- 543
 QY 1032 LPRPIFPNGMNPNTVPCGTEKNTNGMNGHMPSLPDSPTGYIIAVHRKWM-----RT 1086
 DB 544 -PRELL-NGIKD-----NEH-----IVAYRFQMHKGPVK 572
 QY 1087 ELYFLSQKQR-----PSLFQMPBLIVPCTVHTRKQOLYDA-VMIQVRLASPL----- 1133
 DB 573 KLEILHGEQKSSDRGPKCFGTPLV-----TYINKEPLSGTDIATISGLSLPLRVEHMSC 628
 QY 1134 -----PQBEAS-----NHAODCDDSGMYQ-----YPTFLRVVQKDGNSCAWCP 1171
 DB 629 VVNSGNEGHVDPBSSSILSRDTEHDNRELSSLRLRYYSFNLPLESDS----- 681
 QY 1172 WYRCRCCKIDCGEDRAFIQNAIYAVDWDPTALHLRYQTSQERVVDE-HESVQSRAQA 1230
 DB 682 -----VVPNG-----SVTKVLVKWKE-HEKYDSSYLNDLPKVHKNV-LAKKTMQ 725
 QY 1231 EPINLDSCLRAFTSEELGENEMYCKSKTHCLATKKLDLWRLPPLIILHLKRFQFVNG 1290
 DB 726 EGISLFCLEAFIAEPLGPDMMWYCPCKEHRQANKKLDLWLPDLVFLHKKRFTY--S 783
 QY 1291 RWIKSQ--KIVKFPRESFDPFAFLVPRDPALCQHKPLTPQGDSELPRIAREVKYKDAQ 1348
 DB 784 RYFKNKIDTLVNFHIDL-----LSKVYKNEGQ 813
 QY 1349 SSAGEEDVLKSPSSLSANIISSPKSPSSSRKSGTSCPSSKXNSPNSPRLGRSKGR 1408
 DB 814 S----- 814
 QY 1409 LRLPQIGSKNKLSSKENLASKENGAGQ-CELADALSRGHVLGSGQPELVTPQDHEVAL 1468
 DB 815 ----- 814
 QY 1469 ANGELYEHEACGNGYNGQLGNHSEBSTDQREDTRIKPIYNLVYATISCHSGLGGHYV 1528
 DB 815 ---LYLE-----LYAISNHYGGLGGHYT 835
 QY 1529 TYAK-NENCKWYCVNDSSCKELHPDEITDSAYILFYEQGIDYAOPLKPTDGGKADTS 1587
 DB 836 AYAKLMDETKWYFDDSRVSAVNESEIKTSAATVLFYQV-----KSUSETS 882
 QY 1588 SMDED 1592
 DB 883 DMKMD 887
 RESULT 7
 158376
 hypothetical protein unsp - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 31-Dec-1996
 C:Accession: I58376
 R:Gupta, K.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.; Gray, D.A.
 Oncogene 8, 2307-2298, 1993
 A:Title: Unp, a mouse gene related to the tre oncogene.
 A:Reference number: I58376; MUID:9330575; PMID:8336951
 A:Accession: I58376
 A:Status: preliminary; translated from GB/EMBL/DBU
 A:Molecule type: mRNA
 A:Residues: 1-794 <RES>
 A:Cross-references: GB:L00681; NID:G202292; PID:G202293
 C:Genetics:
 A:Gene: unp
 Query Match 11.5%; Score 980.5; DB 2; Length 794;
 Best Local Similarity 25.4%; Pred. No. 6.2e-55;
 Matches 266; Conservative 137; Mismatches 263; Indels 383; Gaps 25;
 663 EEDMLWLYNSNYLTLDDEDHKLEVLKIOEQHLVIEVRNKMSPMEANSFIANSK 712

```
Db 10 ERETRLWNTKMSYVTEQLSKLNTDIQDAGLYQGVLVIEPQNEQDGTWPQSQSKSSTAP 69
Qy 713 -----MSFTAN-----SSKIDR-----HKVPT--- 729
Db 70 SRNFTTSKPSASPYCSVASLIANGDSYNSGMSHSSGVSRSQGSFSAFYNOEAPSPHI 129
Qy 730 EKGATGLSNLGNTCFNMSSIQCVSNQPTQVPIGRHLYELNRPINPIGMKGMAKCYGD 789
Db 130 QPGLCGLNLGNTCFMSALQCLSNAPTLEYFLKDEYEAENRDNFLGKMGGEIAEYAE 189
Qy 790 LVGELNSGTQKVAPEKLRWTIAKAPRNGFOQDSQELLAFILDLGLHEDLNVRHEKPY 849
Db 190 LIQOMSGRDTHTVAPRFKTVQGRFAPQSGYQOQDSQELLAFILDLGLHEDLNVRKCKPY 249
Qy 850 VELKSDGRPDWEAAEAMNHLRRNSIVVDLFGHQLRSQVKCKTCGHSIVRFPFNFPL 909
Db 250 LEPKEPNRGPDAVAXEAMENHLRNDIVVDLFGHQLRSQVKCKTCGHSIVRFPFNFPL 309
Qy 910 SLPLPMDSYNHLBITVILKLD-GTTPVRYGLRLNMDKTYTGLKKQLSDICGLNSEOILLAE 968
Db 310 TLEPLKKORIMEVFLVPADPQCRPIQYRTVPLMGAISDLCEALSKLSGLIAENWVTD 369
Qy 969 VHGSNI-KNPPDONQKVRLSVGFCAFLPVPVPSISASSPTQDTDFSSSPSTNEMFLIT 1027
Db 370 VYNHREHKIFQMDGSLHITPRDDIFVYV-----CNTMDGSECITL- 412
Qy 1028 TNGDLRPIPIPNMGNTVPPCTEKNTNGWNGHMPSLDPSPTGYIIAVHKKMRTTE 1087
Db 413 -----PVYF-----REKK-----SRPSSASSGAV----- 431
Qy 1088 LYFLSSQKNRPSLFGMLPIVPCIVH-TRKOLYDAYWIOVR-----LASPLP- 1135
Db 432 -----LYGQPLLVSPKPKLTLESIIQAVCDRISRVIKQPLDPEFLSPLEPG 479
Qy 1136 -----QASNAHQD-----CDDSMG-----XQYPT 1156
Db 480 ACNGSRSSYEGDEBEHMDHQEQLESEVSGGEDDQGDHSHESAQKVKGPQRHKLFT 539
Qy 1157 LRUVQKDGNSCANCPWVRCRCKIDGEDRAFINAYIADVDPALHRLVQTSGERV 1216
Db 540 FSLV-----NSCGTADINSLATDGL-----LKNRSSTLAIDWDSSTRSLYFDEQSEAC 590
Qy 1217 DEHESVEQSRRAQEPINLDSCLRAFTSBEELGENEMYCSKCKTHCLATKKLDLWRLLP 1276
Db 591 EKHLNSQPQKKKAAVALRECIETFTMETLGEHDPWCTCKKHQOATKFDLWSLEK 550
Qy 1277 ILIHLKRFQFVNGRW-IKQKIVKPRPSFDPSPAFVPRDPALCOHKPLTPQGBELSEP 1335
Db 651 ILVVHLKRFYSY-NRYWRDKLDTVVEFFVRALNMSEF----- 685
Qy 1336 RILAREVKVDAQSSAGEEDVLLSKGSPSLSANIISPKGSPSSSRKSGTSCFSSKNSSP 1395
Db 686 ----- 685
Qy 1396 NSSPRTIGRSKGRLLRFPQIGSKNKLSSKENLDASKENGAGQICELADALSRGHVGGSQ 1455
Db 686 -----VC----- 687
Qy 1456 PELVTPQDHEVALANGFLYEACNGSYNGQLGNHSEEDSTDQREDTRIYP-IYNLYA 1514
Db 688 -----DRSAPYVVDLIA 700
Qy 1515 ISCHSGILGCGHYVYAKNP-NCKWYCYNDSSCKELHPDEIDTDSAYILFYEQQGIDYAAQ 1573
Db 701 VSNHYGAMGVGHYTAVALKRLNGKYYVFDSSVSLASEDQIVTKAAYVLFYQRRDECSS 760
Qy 1574 -----FLPKTDGKKMADTSSMDEDFSDY 1597
Db 761 TSSLGSPFGDGGVKLSSSHQGMGDEEAY 789
```

RESULT 8

```
T04192
hypothetical protein T4F9.30 - Arabidopsis thaliana
N:Alternate names: protein T4F9.30
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
C:Accession: T04192
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15260
A:Accession: T04192
A:Molecule type: DNA
A:Residues: 1-928 <BEV>
A:Cross-references: EMBL:AL049523
A:Experimental source: cultivar Columbia; BAC clone T4F9
C:Genetics:
A:Map position: 4
A:Interons: 51/2; 122/2; 161/3; 181/3; 214/3; 328/1; 401/3; 458/3; 544/3; 596/2; 774/2;
A:Note: T4F9.30
```

```
Query Match 11.3%; Score 967.5; DB 2; Length 928;
Best Local Similarity 24.3%; Pred. No. 5.5e-54;
Matches 320; Conservative 153; Mismatches 352; Indels 493; Gaps 38;
Qy 359 VCHIVLGLRATPEEEQIIRGHLERESRYGLOAGHNWFLISQWQWQWKEYVYKIDANPV 418
Db 14 VCDL-----PFTPEEEKRIV-SELTSEEDNLKQGNLYFVISKRWYTSWQYVENSANE- 66
Qy 419 VIEPSSVLNGGKYSFGTAHPMEQVEDRIGSSLSVYNTTEKEFSDNISTASEASETAGSG 478
Db 67 -----CSTGESSEA----- 75
Qy 479 FLYSATPGADVCFARQHTSDNNNOCLLGANGNILLHNPQKPGALDNPQVLPQEPKAT 538
Db 76 -----PRPGPIDNHDIISDS----- 91
Qy 539 SLTLEGGRKRLTPOLIHGRDYEMVPEPWRALYHWYGANLALPRPVTKNSKTDIPELELF 598
Db 92 --DINDQLREL-LVSGEDYVLVPQVWKLVECG- -----PIE-- 128
Qy 599 PRYLLFRQAPATQTSNIWNMGVPSNAPLKRVLAYTGCFSRMOTIKETHEIYLSQR 658
Db 129 -RKLIC--QGFYTRSYSVEVY-----FLCLMLTDGRDESRTASIRELYEKVCAM 174
Qy 659 LRKEEDMRWLNSNYLTLLDDEHK-LLEYLKIQDBQLHVIETVRKMSWPEEMSFA 717
Db 175 TGVQVEKAHWDYDKRKNGLLDPLSYKLEBSLHMDQDILVEVDGLSSSSQSSASSTG 234
Qy 718 NSSKIDRHKVPTKX----- 731
Db 235 NELAL-----VPLEPSRGIVTIAGPTLSNGHSTTSNPSLPFRITSEDDGRDSLILKGE 290
Qy 732 --GATGLSNLGNTCFNMSSIQCVSNQPTQVPIGRHLYE-----LNRTNP 776
Db 291 KGGLAGLSNLTGNTCFMSALQCLAHPTPIVEYFLQDYMLDTPYFYLWVLTICILNILT 350
Qy 777 IGMKHVAKCYGDIVQVLSGTQKNVAPLKLRTIARYAPRFGNGFOQDSQELLAFILDG 836
Db 351 LNFQGEIAAFGDLKKLWSSGRNAVAPRAFKTLARFAPQPSGYNCHDSQELLAFILDG 410
Qy 837 LHEDLNVRHEKPYVELKSDGRPDWEVAAEAMNHLRRNSIVVDLFGHQLRSQVKCKTC 896
Db 411 LHEDLNVRKPKPYIELKSDSRPDDEVAEELWNYKARNDSVIDVCQGGKSTLVCPCVC 470
Qy 897 GHISVRDPDFNLSPLPMDSYNHLITVILKDG-TPVRYGLRLNMDKTYTGLKKQLSD 955
Db 471 GKISITDFPFWYLSVPLPSTLTSMTITVFCGSRPLPMEYTVIVPKQSIRDLITALGT 530
Qy 956 LCGL-NSQEIILAEVHGSNKNFPQDNQKVRLSVGFCAFLPVPVPSISASSPTQDTDF 1014
Db 531 ACCLAEDESILLAEVYDHKIFRY-----FEIPLD-----SLSAIKDDEH 569
Qy 1015 SSSPSTNEMFLITNGDLRPIPIPNMGNTVPPCTEKNTNGWNGHMPSLDPSPT 1072
```


802 VLVHLKRFYSRSMKLETFVNFPIHDLTKYV----- 837
1337 ILAREVKVDAQSSAGEEDVLLSKSPSSLSANIISPKGSPSSSRKSGTSCPSSKNSPN 1396
838 ----- 837
1397 SSPRTLGRKGRRLPQIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLLGGSP 1456
838 -----ANKNL-----SQP 845
1457 ELVTPQDHEVALANGFLYEACGNGYNGQLGNHSEEDTDQREDTRIKPIYNYLAIS 1516
846 QL-----YLYALT 854
1517 CHSGLGGGHVYTYAK-NPNCKWYCYNDSSCKELHPDEIDTDSAVILFYEQQ 1567
855 NHYGGMGSGHYTAHKLDDSRWYNFDDSHISHINEDDVKSGAAVLYFYRRK 906

RESULT 10
T41006
ubiquitin carboxyl-terminal hydrolase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T41006
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
Submitted to the EMBL Data Library, June 1998
A:Reference number: Z21963
A:Accession: T41006
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-979 <WOO>
A:Cross-references: EMBL:AL023776; FIDN:CAAL9303.1; GSPDB:GN00067; SPDB:SPCC1494.05c
A:Experimental source: strain 972h; cosmid ci494
C:Genetics:
A:Gene: SPDB:SPCC1494.05c
A:Map position: 2

Query Match 9.2%; Score 787.5; DB 2; Length 979;
Best Local Similarity 22.9%; Pred. No. 2.5e-42;
Matches 283; Conservative 162; Mismatches 414; Indels 377; Gaps 37;

432 SFGTAHPMEQVEDRIGSSLSVYNTTTEEFKFSNISTASEASETAGSGFLYSATPGADVCF 491
19 SLSEESQSSNMDDISQKISLGDASE--ISKNLPSIAEQKQLIG----- 61
492 ARQHTSDNNQCCLIGANGNILLHNPQK-----PGADNOLPVTQBPVKATS 539
62 ----ELVNNQPELELGQVDNYILSYSWYERLCSYLAEDGFFGPDVEDIA----- 108
540 LTLGGRLKRTPOLIRGRDYEMVPEVWRALYHWYG-ANLALPRVINKSKTDIDPEL--E 596
109 -DLETGTLK--PDQBEIDFTIISRDVLLVWVGLKGPPEFRTVMLGSESHPLVVE 165
597 LPPR-VLLFLRQOPARTQCSNIWVMGNVSPNAPLKRVLAYTCGFRMOTIKEIHYEL 655
166 VYPIFSLTLLSTNAVANES-----HKPKKISL-----SKSTLEDLLSGV 207
656 SQRLRKEEDMRMLVNSNYLTLLDDEHKLKYLKIQDECHLVIEVKNKM----- 707
208 KYTSLSPSQFRLWRVDTQPLHRTIDPS---SFKINSKEIIDLEKSKILVELGMDS 264
708 -----SWPEE-----MSFI-----ANSKIDRHKVPTKEGATGLSNLGNTCF 744
265 CSLVAECMINETWVDRALRLQFLIQORNQSSNEEQKEKVP---GTCSGLSNLGNICY 321
745 MSSICCVNTQPLTYQFYISGRHLYELARTNPIGMKGMKACVGLVDELWSTGTQNVAP 804
322 MNSALQCLTHRELDRFFTSDEWKQVNESNPLMGQGVQVAFSLIKLSLSPHSSFPAP 381
805 LKLRWTIAYAPRFNGFOQDSQELIAFLDGLHEDLNRVHEKPYV---ELKDSGRPDW 861
382 RQFKATIGFNHSGFLGYGQDSQEFALFLDGLHEDLNRIYQKPYTSKPDLYEVDDEKIK 441

QY 862 EVAAEAWNDHLRNRSIVVDLFFGQLRSOVKCKTCGHSVREDPFNFELSLPLPMDSTMHL 921
DB 442 NTAEECWRLKRLNDSLVLDLFOGMYRSTRLLVCPVNTVSTITDFPFMDLTLLPLPVKQWWSH 501
QY 922 EITVIKLD-GTTPVRYIGRLNMDE-----KYTLKKQLSDCLGNSEQIILAEVHGSN 973
DB 502 TWTPIADTNTLPLAIVLVLESKAATIEDLVKVAESKGCSDY-----RKILVETIYKGR 556
QY 974 IKNF-PODNOKVRLSVSG-----FLCAFEIP-----VPVSPISASSPTQTD-FSSSP 1018
DB 557 FYRFLTQLSKSLMEISEDEIYLYELERFEDYDGSDDILVPVYTHISDDSTNSANSYMSSR 616
QY 1019 STNEMFTL-TTNGDLPRPIPIPNMGMENTVVPCTEKNFTN-----GMVNGHMPSLPDSPPF 1072
DB 617 DFGHPFVLQSDNEVTDASFISEKIKLYQQFTTLKLNKNIDLSLELGH-----EDEQV 672
QY 1073 TGYIIAHRKMRTELTYLSSQNR-----PSLFGMPLI-VPCTVTRKKDLYDAVWQVS 1127
DB 673 QKGLDVMDDHSQTPLFEMRVFHDREFEIKPTGNWNGVSNLPLLTERDKKD----- 722
QY 1128 RLASPLPQEQASHAODCDDSMGVQVPTLRLVVKDGNSCAWCPWYRFCRGCKIDCGE-D 1186
DB 723 -LESTVDPLDAHSIESEDSEPKDVAFGSYPEPSKSNEM-----TKLTAKEND 769
QY 1187 RAFIGNAYIAVDWDPTALHLRYQT-----SQERVVDEHESVEQSRRQAEP 1232
DB 770 RLLTQGLLVCEWPEKSYQVFSVAPSSPQMGSLMLESKTLSDKKDSDSR-----T 824
QY 1233 INLDSCLRAFTSEELGENEMYCSKTHCLATKLDLWRLPPLIILHKKRFQVNGRW 1292
DB 825 ITLNDCLDEFEKTEQUGEDFPWYCPCKEFRQASKQMEIWRCEPILIFHLKRFSSRRFR 884
QY 1293 IKSQIKVFPRESFDSAFIVPRDPALCOHKPLTPQGDSELPRIAREVKVYDAQSSAG 1352
DB 885 DKIDDLVEFP----- 894
QY 1353 EEDVLLSKSPSSLSANIISPKGSPSSSRKSGTSCPSSKNSPNSSPRTLGRSKGRRLP 1412
DB 895 ----- 894
QY 1413 QIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLLGSGQPELVTPQDHEVALANGF 1472
DB 895 -----INDLDMRTGSKYKLE----- 911
QY 1473 LYHEACGNGYNGQLGNHSEEDTDQREDTRIKPIYNYLAISCHSGLGGGHVYTYAK 1532
DB 912 -----KENPKL--IYELYAVDNHYGGLGGGHYTAFAK 941
QY 1533 NP-NCKWYCYNDSSCKELHPDEIDTDSAVILFYEQQ 1567
DB 942 NPDNGQPYCFDDSRVTPVCPPEETVTSAAVLLFYRRK 977

RESULT 11

T33734
hypothetical protein H34C03.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33734
R:Kalicki, J.; Gibson, A.
Submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid H34C03.
A:Reference number: Z21394
A:Accession: T33734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-900 <KAL>
A:Cross-references: EMBL:AF100662; FIDN:AAC68977.1; GSPDB:GN00022; CESP:H34C03.2
A:Experimental source: strain Bristol N2; clone H34C03
C:Genetics:
A:Gene: CESP:H34C03.2
A:Map position: 4

A; Introns: 25/3; 85/1; 148/3; 170/1; 216/3; 247/2; 326/3; 506/3; 556/2; 816/2; 864/1

Query Match 8.8%; Score 756.5; DB 2; Length 900;
Best Local Similarity 21.1%; Pred. No. 2.2e-40;
Matches 276; Conservative 153; Mismatches 339; Indels 539

Qy	1316	DPALCQHKPLTPQGDSEPRILAREVKVKDAQSSAGEEDVLTKSPSSLSANIISSPKG	1375
Db	756	-----	755
Qy	1376	SPSSSRKSGTSCPSSKNSSPNSPRTLGRSKGRRLPQIGSKNKLSSKENLDASKENGA	1435
Db	756	-----SREKLIW-----	762
Qy	1436	GOICELADALSRGHVLGSGSOPELVTPQDHEVALANGELYEHAECAGNGYNGQOLGNHSEED	1495
Db	763	-----EVVIP-----VRGL-----	772
Qy	1496	STDDQREDTRIKPIYNIYLAISCHSGILGGGHVYTYAKNP-NCKWYCVNDSSCKELHP--D	1552
Db	773	LTDKVANPNHEKAYDLLAVSRHYGSLGGHYTAIGNDREQKWPFDNUSANPTFPAPAE	832
Qy	1553	EIDTDSAYILFYEQQIDYVAQFLPKTDGKKMADTSSMDDEDSPEDYKK	1599
Db	833	FYESSDFYILVYRRKLD-----SNGKPI-EPSAVSESLSHSPKRR	871

RESULT 12

RESULI 12
S46636
hypothetical protein YJL197w - yeast (*Saccharomyces cerevisiae*)
N/Alternate names: hypothetical protein J0340
C/Species: *Saccharomyces cerevisiae*
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 29-Oct-1999
C/Accession: S46636; S56984
R/Purnelle, B.; Coster, F.; Goffeau, A.
Yeast 10, 1235-1249, 1994
A/Title: The sequence of a 36 kb segment on the left arm of yeast chromosome X identifies gene ACO1 and two homologues to chromosome III genes.
A/Reference number: S46621; MUID:95274326; PMID:7754713
A/Accession: S46636
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-1254 <PUR>
A/Cross-references: EMBL:X77688; NID:gl183992; PIDN:CAA54762.1; PID:g547597
R/Purnelle, B.; Coster, F.; Goffeau, A.
submitted to the Protein Sequence Database, September 1995
A/Reference number: S56977
A/Accession: S56984
A/Molecule type: DNA
A/Residues: 1-1254 <PUW>
A/Cross-references: EMBL:Z49472; NID:g1008412; PIDN:CAA89492.1; PID:g1008413; MIPS:YJL1
C/Genetics:
A/Gene: SGD:UBP12
A/Cross-references: SGD:S0003733; MIPS:YJL197w
A/Map position: 10L

Query Match	8.6%	Score 738.5;	DB 2;
Best Local Similarity	21.8%;	Pred. NO. 5.3e-39;	Length 1254;
Matches 296;	Conservative 15;	Mismatches 384;	Indels 521;
			Gaps 47;

QY	445	DRIGSLSVNTTEKFSDNISTAS-----EASTAGSGFIYSATPGADVCFARHNTS	498
Db	29	DLRGVDL--MNYLDK--DEIKOESVPVSREIETEDSASAVSFAGANEIIAEPHAAS	84
QY	499	DNNNOCLLGANG---NILLHLNPKGAIHQPLVTQPEPVKATSLT-LEGGRLKRTPOL	553
Db	85	ETN-----LGTNGQDGRNVL-----EQORDVARLISENKETOEGDKVICIVPKV	129
QY	554	IHG-----RDYE-----MVPEPVRALYHYHYGANL	578
Db	130	WTKDFDPDVTDPIGINTXMI CRDFENFVLBYNRCPYLSIAEPVNFUSEIYGMTS	189
QY	579	ALPRPVINKSTDDIPELEFPRY-----LLFLRQQPATRTQQ-----SNINWNMGVPS	627
Db	190	G-SYPVTVNLVINQTTGBLETYNKWFFRLHYLTKEQGRKRHHQDDDSIMYLSM----	243
QY	628	PNAPLKRVLAYTCFSRMOTIKEHE-----YLSQRILRIKEEDMRJLW-----LYNS	673

Query Match		8.8%	Score 756.5	DB 2	Length 900
Best Local Similarity		21.1%	Pred. NO. 2.2e-40		
Matches		276	Conservative 153	Mismatches 339	Indels 539
		Gaps	40		
A; Introns: 25/3; 85/1; 148/3; 170/1; 216/3; 247/2; 326/3; 506/3; 556/2; 816/2;					
QY	370	TPEEQSIIRGLWLESRESYGLQAGHNFIISQWQWQWKKYKYKDANFVVIPESSVLNG	429		
Db	27	TSRE--VVKLLNEVEAEENEMWYLSQKWN	60		
QY	430	KYSGTAAHMEQVEDRIGSSLSYVNTTEEFSDNISTASEASETAGSGFLYSATPGADV	489		
Db	61	--SLIKAVRDFGVDCGPI	77		
QY	490	CFARQHTSDNNQCLLGANGNILLHNPQKGAIDNQPLVTQBPVKATSLTLEGRRLK	549		
Db	78	--DNSQISECRNCSFFL	100		
QY	550	TPOLHGRDYEMVPEVVRALYHWY	604		
Db	101	--DYTPVPEILFKELERFVHDDRRDYIQRVVRKQKLCI	149		
QY	605	LQOPATRTQQSNINWNGNVSPNAPLKVLYATGCFSRMOTIKEIHYLSQRLRIKEE	664		
Db	150	--TSANRDIKADLVRPD	185		
QY	665	D-MRLWLVNSNYLLDDEDHKLEYLKIODEQHLVEVN	710		
Db	186	DNIRFVWNGNY-ELIDTASIDSY	242		
QY	711	EMWSFIANSKIDRHKVP	761		
Db	243	QSSSHGGNGIGISNSYNSPSTGVVTRTPCAGSLSLNGTICPMASAFQCLSNMPLREY	302		
QY	762	FTSGRHLYELANTNPIGMKHMAKCYGDLVQELWSGTQKNVAPLKRWTIAKYAPRNGP	821		
Db	303	FIANYQNDINEDPLGTHGLHMAVGLMKMGWSGDYASINPRKPKSIIGOFAPRPNFY	362		
QY	822	QQDSQELLAFLLDGLHEDLNVRHEKPYVELKQSDGR	880		
Db	363	SQDAHELWAYVLDGLHEDLNRIKKKPYIEDSDDEDAKLPEAEYAAKSWEYKORNDISIIV	422		
QY	881	DLFHGLQBSQVCKCTCGHISVRPDPFFLSLIPMDOSYMHLEITVIK	928		
Db	423	DTLHQLKSTLPCVCEKISIKPDPFGLSLPFPKEQIVKQTVIWMELTRKWFSLGI	482		
QY	929	DGTPVRYGLRLNMDKTYGLKKQL	969		
Db	483	TDTTTVEAERLMLDKLQPEQHHFVFFHVPFSQYHDDITMRPGRVMPGREVYVAEVE	542		
QY	970	HGSNINPQDNQKRVLSVSGFLCAPEIPVPSFISAGSGTQDFSSSPSTNEMFTLTN	1029		
Db	543	HDLOVKG--TRLFA	564		
QY	1030	GDLPRPIFGMNTVVPCTGTEKNFTNGMVGHPSLPDSPTGYIIVHRCWMTLEY	1089		
Db	565	ASLPMIYSLPTGY	595		
QY	1090	FUS-----SQKNRPSLFGMLPIVCTVHTRKKLYDAVVIQV	1135		
Db	596	FINPRKVAELGGNSNDDDSMSG	645		
QY	1136	QEASHNAQCCDSMGYQYFFTLRVQKQNSCAWCWFYRCRCKIDCGEDRAFTGNAYI	1195		
Db	646	-----TDDPI-----PF-----PDENS-----	667		
QY	1196	AYDWDPTALHRYQTSQERVVDEHESVEQRRQAQAPINLDSCLRAFTSEBELGENEMY	1255		
Db	668	VFOKDLXIFNQYKGNELIEREMTVSTR	720		
QY	1256	CSKCKTHCLATKLDLWLPILLIHLKRFQFNGRWIKSQKIVKPPRESFDPGSAFLVPR	1315		
Db	721	CPOCKXGPRATKQALAKWLPBILTLHLKRFQY--KW	755		

Db 244 -----SALNLVRLVLEKSMNLFPEKADHLVDNAVDFKIFWVSEGSDIATD 288
QY 674 ENYLITLDD--EDHKLVELKI-----ODEQLVIEVR--NKDMSWP 710
Db 289 SNVSIFLNSSEYITPLQFLELPIKLLIPDMFENRLDKITSNPSDLVIEIKPIEGNHWP 348
QY 711 EWSFIANSKIDRHKVPTEKATGLSNLGNNTCFWNSSIQCVSNTPQITQVTFISGRHLYE 770
Db 349 -----SNFYANKLEPASOTGLVLNLTGNTCYNNSALQCLVHIPQLRDYFLVDGYDE 400
QY 771 LNRNTPIGMKHMAKCYGLVQELM-----SGTQKXVA--PLKLRWTIAKYAPRNGFOOQ 824
Db 401 INEENPLGVHGVARAFSDLVOKLFQNRMSIMORNAAPPSMFKSTIGHFMSFGYMQQ 460
QY 825 DSQELAFLLDGLHEDLNVRVHEKPYVE-----LKSDGRPDWEY-----AAEAWNDHLNRNS 877
Db 461 DSQEFLLDLSLHEDLNRIIKKETKPSLSPGDDVNDVNVKGLADTWEMHLKNC 520
QY 878 IYVDLPHGLRQSVCKTKCGHISVRDPNPLSLPLPMDSYMHLEITVILKIDGTPVRYG 937
Db 521 VITDLPVGMKSTLYCPECONVSIITPDYNDVTLPLPVDTVMDTKIKIPFN-SPPLLE 579
QY 938 LRLNMDKTYTGLKKQLSDLCGLNSQILLAEVHSGNIKNFPQDNOKVRLSVSGFLCAPEI 997
Db 580 VELSKSSTYMDLKNVYGMKGLDPTLFCGEIFSNQI-----YVNVYESTESNAQFLTLOEL 635
QY 998 PVPVSPISASSPTQDFSSPSSTNE-----MFTLTNGDL 1032
Db 636 IRPADDI-----FVELPVTNDNEVIVPVLNTRIEBKYNAMLFVGFPPFITKEDEL 687
QY 1033 PRPIFIPNGMPTV-----PCQTEKNFTNGMVNGHMPSL-----PDSPTGTG-- 1075
Db 688 NNPGALRMKLQNRVHLSGVIYFPFVGNRTDFADA-----FELLVVKYDPVFEQYKD 742
QY 1076 -----IAVHRKMRTELYFLSSQKRPSPFLGMPLI----- 1106
Db 743 ILQYTSIKVTRDKSGFFSILSVK-----EQFASNNRTGPN-FWTFIQLNLDRATD 796
QY 1107 VPCTVHTRKQILY-----DAWIOV-----SELASP-----LP 1134
Db 797 IDKLEDDVVDIYNTSSLDVCAEGLMVQDDDEGTGSEAKNFPFGSGDDEENKETVT 856
QY 1135 POEASNAHQDCDDSMGY-----QVFTLRVQKDN 1165
Db 857 NNENVNNTDRDEDMELTDVDEADASTEPELTDKPEALDKIKSITSTPFALLSN-----N 912
QY 1166 SCACWPTFRFCGCKIDCGEDRAPIGNAYIADVDPALHRYQTSQBRVVDHESVEQS 1225
Db 913 DIIVCEW-----SELGSNEAFSDDK--IYNWENPA-----TLPNKL-ENAKLERS 955
QY 1226 RRAQAEPINLDSCLRAFTSEELGENEMYCYCKKTHCLATKLDLWLPPIILHILKRF 1285
Db 956 -NAKERTITLDDCLQLFSKPEILGLTDSWYCTCHEHQATKQLQWNTPDILLIHLKRF 1014
QY 1286 QFNGRWIKSQIKVPPRESFPFPAFLVRDPALCOHKPLTPQGBELSEPRILAREVKV 1345
Db 1015 ESQSFSDXIDATVAFPIITDLSRYVVKD----- 1045
QY 1346 DAQSSAGEEDVLLSKSPSSLSANISSPKGSPSSRKSGTSCPSKSNPSNPSRTLGSR 1405
Db 1046 -----DRGL----- 1050
QY 1406 KGRLLRPIQSGKNKLSSKENLDASKENGAGQICELADALSRGHVGLGSGQPELVTPDQHE 1465
Db 1051 ----- 1050
QY 1466 VALANGFLYEACNGYNGQLGNHSEBSTDQDRETRIKPINLYALISCHSILGGG 1525
Db 1051 -----YDLVAVDNHYGGLGG 1067
QY 1526 HYVTYAKN-PNCKWYCYNDSSCKELHPDEIDTDSAYILFY 1564
Db 1068 HVTAYVGNFADKNWYFDDSRVTETAPENS IAGSAYILFY 1107

RESULT 13

T41085

probable ubiquitin carboxyl-terminal hydrolase - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T41085

R/McDougall, R.C.; Rajandream, M.A.; Batrell, B.G.; Zimmermann, W.; Wambutt, R.
submitted to the EMBL Data Library, August 1999

A/Reference number: Z21822

A/Accession: T41085

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-849 <MCD>

A/Cross-references: EMBL:AL109957; PIDN: CAB53084.1; GSPDB: GNC00068; SPDB: SPCC16A11.12C

A/Experimental source: strain 972h-; cosmid c16A11

C/Genetics:

A/Gene: SPDB: SPCC16A11.12C

A/Map position: 3

A/Introns: 107/2

Query Match

8.1%; Score 693.5; DB 2; Length 849;

Best Local Similarity 21.3%; Pred. No. 2.3e-36;

Matches 233; Conservative 160; Mismatches 321; Indels 379; Gaps 32;

QY 534 PVKATSLTLEGGRLKRTPOLIHRDYMVPVPRVRYALYHYG-ANLALPRPV--IKNSKT 590

Db 70 PITQWELLDKKNELKHS--LEESIDYSIVSASLWHLVEMFGLGLEGLAERKVLVGLAAE 127

QY 591 DIPELELP-RVLLFLRQOPATRTQOSNIWVMGNVP-SP-----NAPLRVLAAYTCPSR 644

Db 128 QKPFVDIYFINFLVHLFDP-----INGENTSPLYQIDEPYHSDPEYAFPSR 177

QY 645 MOTIKIHYLSORLKE-EDMLWLNVSEN---YLTLDDDEHKL-----EY----- 689

Db 178 SDTLRLSLYQVMEAFQISDGSFRLWLNKSNLSRFVSLSEFNQOPATALLSEYAVCM 237

QY 690 ---LKTQDQHLVIVRNKDMWPBEMSFANSKIDRHKVPTEKATGLSNLGNTCFNV 746

Db 238 IFIDIDADGS-LLELQHPENGWLSL-----SITQEQLNTINKEIGLCGLVNLGNSCYN 291

QY 747 SSIQCYSNTQPTQVTFISGRHLYELNRTNPIGMKGMKACYGDLVQEL-WSGTQKNVAPL 805

Db 292 SALQCMHTEHETKPLSDSYEKDINYNNPLGMGKVALSYASLLKMIHTADMHSVSPS 351

QY 806 KLRWTIAKYAPRNGPQQDSQELAFLLDGLHEDLNVRVHEKPYVELKSDGCRPD----- 860

Db 352 SPKFIICEFTVPSGYRQDSQEFIAFFLDGLHEDLNRIQIKPYFE-----RPDLFDSH 405

QY 861 ---WVVAEANDNHLRNRISIVVDLPHQLBSQVKCKTCGHSVRPDPFNLSLPLMD 916

Db 406 PLHVQRVANQCDIHRKNDISIIVQLFQMYKSTLECSICYOKSTAFDPFMTLPLPTS 465

QY 917 SYMHEITVILKIDGT-TPVRYGLRLNMDKYLTKKKQLSD---LCGLNSEQILLAEVHGS 972

Db 466 AKRHRKVVVPFGTQSPVELYLELLMESTVIQMFQATEKLQKMGLEGELTACDIYRG 525

QY 973 NIKNFPQDNQKRLSVSGFLCAFEIPVPSIPASSTCTDFSSSPSTNEMFTLTNGDL 1032

Db 526 KVKYLNKNDKTSKKT----- 541

QY 1033 PRPIFIPNGMPTVPCGTEKNFTNGMVNG-HMPSLPDSPTGTGYIIAVHRKMRTELYEL 1091

Db 542 -----HKWDHVLYGSTANGLTIPIVHGCKRPAMFG----- 572

QY 1092 SSQKNRPSLFGHPLIVCFVTHRKD-----LYDAVWIOVSRSLAPLPQ 1136

Db 573 SYGSN--DVFGFPL-----QLNYSRNLVNDLVKSEIVELYRYAGIDVAIGTLQLGKEM 626

QY 1137 EASNAHQDCDDSMGYOYPTLVRVQKDGNSCAWCPWYRCRCKIDCGEDRAPIGNAYIA 1196

Db 627 ESKAGWELIKIEVK---RFEIVBEE-----EIV 653

QY 1197 VMDPTALHLYQTSQ-ERVVDHESVEQSRRAQAPINLDSCLRAFTGEEELGENEMY 1255
Db 654 ID-DKTVIMCLWNDQOYEKLFYNCEWIFEKIQHFMESITLEDCLLEFSKPEQLDLODSWY 712
QY 1256 CSKCKTHCLATKLDLWRLPILIIHLKRFQVNG---RWIKQKIVKPPRSPFDSAPL 1312
Db 713 CPCKAPRPATKLEIWRLEPKILVILHNRFSHGGLRRRRKRDVAVPVFDLNLKQEL 772
QY 1313 VPRDPALCQKPLTPQDELSEPRILAREVKVDAOSSAGEEDVLLSKSPSSISANI 1372
Db 773 SP----- 774
QY 1373 PKGSPSSSRKSGTSCPSKSNKSNPSRPTLGRSGKRLRLPQIGSKNKLSSKENLDASKE 1432
Db 775 ----- 774
QY 1433 NGAGQICELADALSRGHVGLGSOPELVTPQDHEVALANGFLYEHEACNGYNGQLGNHS 1492
Db 775 -----FXDHEWLSQKS----- 787
QY 1493 EEDSTDQREDTRIKPIYNYLAISCHSGILGGHYVYAKNPCK-WYCYNDSSCKELHP 1551
Db 788 -----MLDYAVDNHHGFMNSGHYAYARDASSQTFKFDTAICEIDP 832
QY 1552 DEITDTSAYILFY 1564
Db 833 EDIVTSSAYILFY 845

RESULT 14
T20528
hypothetical protein F07A11.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20528
R:Palmer, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19287
A:Accession: T20528
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1095 <WIL>
A:Cross-references: EMBL:Z66511; PIDN:CAA91317.1; GSPDB:GN000020; CESP:F07A11.4
A:Experimental source: Clone F07A11
C:Genetics:
A:Gene: CESP.F07A11.4
A:Map position: 2
A:Introns: 32/2; 202/2; 404/3; 587/3; 794/3; 964/1; 1004/3; 1080/2

Query Match 7.9%; Score 675; DB 2; Length 1095;
Best Local Similarity 28.8%; Pred. No. 5.4e-35;
Matches 197; Conservative 81; Mismatches 223; Indels 182; Gaps 18;

QY 723 DRHKV-PTKGGATGLSNLQNTCFMNSSICQVNTQPLTQYFTISGRHLYELNRTNPIGMKG 781
Db 411 DREKVCWVEQVGTGLRNIGNTCFMNAVLQMLVNNIELREYFLRNHYQPEINETNPLSSEG 470
QY 782 HWAKCYGLVQBLMSGTKQKVAPLKRWIAKYAPRNGFOODSQBLAFLLDGLHEDL 841
Db 471 RLAKAFADFMHQWSHQKAIQBEPTQKNIVABKASQFANFAQDAHEFLSLDGLHEDV 530
QY 842 NRVHEKPYVELKDSQCRPDWEAAAWNDHLRRNRNSIVVDLPHGQLRSQVCKTKCGHLSV 901
Db 531 NRVKKKPLGTGVESHGRHDLVDVNEAKNKHILRNDISIFDLPHGQLKSHVQPCNDRVSI 590
QY 902 RDPDFNLSLPLPMSYMHLEITVILKDTTPRVYGLRLNMDEKYTEGLKQLSLDCLGNS 961
Db 591 TFDPPFVLPVPPFKN-----KQSTDLI----- 612
QY 962 EQILLAEVHGSNIKFPQDNQ-----KYRLSVSGFLCAF-----EIPVPVSPISASP 1009
Db 613 -----FWPLENQTKPKLTVSYSTEGTVADFLSVSVSETVRVPTNLRACEA 658

QY 1010 TOTDF-----SSPSTNEMFTLTNGD-----LPRPIFIPNGMPT-- 1045
Db 659 ICHKFDKIYSLDMKVIDIVSPDLRFVFTHDEMFEFNEIVILHVLORELY-----RKNTKH 714
QY 1046 -----VVPQGTBKNFTNGMVGNGHMPSLPDSPTGYIIAIVHRKMMR 1085
Db 715 ACHECGNSKLKACBECYDAVYCSKEQVANWSTGGH-----R 753
QY 1086 TELYFLSQKRRPS-LFGMPLIVPCTVHTTKKOLYDAWIVQVSLASPLPPQBSNHAQ- 1143
Db 754 EE-----CSKRPSETVGHPLIVSL--PRSLTYQHLVRLVLEAKSRHTPPGESSSSSV 805
QY 1144 -----DCDDSMGYQYPTLRRVQKDGNSCAWCFWYFCRCKTDCGEDRAFINAY 1194
Db 806 SSTPRRSVAAPRTKQKMFIRKLSGQND-----FGSHSISDDEACQLESQGGY 857
QY 1195 IAVNDPTALHLRYQTSQERVVDEHESVEQSRRAQAPIN-----LDSCIRA 1241
Db 858 VSVNWINQRNGKPYITTIENR-KQIDVDVEKSRQWNRNMNGYVKNNSNETNPHLTQMLDL 916
QY 1242 FTSEELGENEMYCYCKTHCLATKLDLWRLPILIIHLKRF-----QFVNGRWIK 1294
Db 917 FSETERLKPEESWTCTCKENVEATKKQLYRUPPVLIILQKRPVYTAFTYQSSMAERSK 976
QY 1295 SQKIVKFPRESFDPFAFLVPRDP 1317
Db 977 DTRSVYPLESLDMSPLAETSP 999

RESULT 15
S39344
deubiquitinating enzyme SSV7 - yeast (Saccharomyces cerevisiae)
N:Alternate names: deubiquitinating enzyme DOA4; protein YBR926; protein
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 20-Jun-2000
C:Accession: S39344; S49824; S54053; S58849; S61758; S30807; S67885
R:Papa, F.R.; Hochstrasser, M.
Nature 366, 313-319, 1993
A:Title: The yeast DOA4 gene encodes a deubiquitinating enzyme related to a product of
A:Reference number: S39344; MUID:94067315; PMID:8247125
A:Accession: S39344
A:Molecule type: DNA
A:Residues: 1-926 <PAP>
A:Cross-references: GB:U02518; NID:G408456; PIDN:AAC48915.1; PID:G408457
R:Richards, C.; Harris, D.E.
submitted to the EMBL Data Library, November 1994
A:Reference number: S49823
A:Accession: S49824
A:Molecule type: DNA
A:Residues: 1-926 <RIC>
A:Cross-references: EMBL:Z46796; NID:G577794; PIDN:CAA86791.1; PID:G577796
R:Hunt, S.; Bowman, S.; Harris, D.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54031
A:Accession: S54053
A:Molecule type: DNA
A:Residues: 1-926 <HUN>
A:Cross-references: EMBL:Z49209; NID:G798897; PIDN:CAA89098.1; PID:G798920
R:Brandt, P.; Otto, B.; Ramlow, S.; Blocker, H.
submitted to the EMBL Data Library, January 1995
A:Reference number: S58832
A:Accession: S58849
A:Molecule type: DNA
A:Residues: 36-926 <BRA>
A:Cross-references: EMBL:X84162; NID:G706817; PIDN:CAA58985.1; PID:G706835
R:Brandt, P.; Ramlow, S.; Otto, B.; Blocker, H.
Yeast 12, 85-90, 1996
A:Title: Nucleotide sequence analysis of a 32,500 bp region of the right arm of Sacchar
A:Reference number: S61741; MUID:96381250; PMID:8789263
A:Accession: S61758
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:32:31 ; Search time 18 seconds
(without alignments)
4190.602 Million cell updates/sec

Title: US-09-888-615-73
Perfect score: 8563
Sequence: 1 MGAKESRIGFVSYEALRRV.....DTSSMDEDFSDYKVKYCVLQ 1604

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4531	52.9	1089	1 UBP6 HUMAN	P35125 homo sapien
2	1087.5	12.7	981	1 UBP6 MOUSE	Q815h1 mus musculus
3	1087	12.7	963	1 UBP4 HUMAN	Q13107 homo sapien
4	1086	12.7	962	1 UBP4 MOUSE	P35123 mus musculus
5	1085.5	12.7	981	1 UBP7 HUMAN	Q944e8 homo sapien
6	903.5	10.6	690	1 UBP6 HUMAN	P51784 mus sapien
7	902.5	10.5	699	1 UBP6 MOUSE	Q99k46 mus musculus
8	787.5	9.2	979	1 UBP6 SCHPO	O60079 schizosacch
9	738.5	8.6	1254	1 UBP6 YEAST	P39538 saccharomyc
10	693.5	8.1	849	1 UBP1 SCHPO	O8um5 schizosacch
11	665	7.8	1371	1 UBP7 HUMAN	Q94966 homo sapien
12	498	5.8	1118	1 UBP8 HUMAN	P40818 homo sapien
13	446	5.2	357	1 UBP2 CHICK	O57429 gallus gall
14	420	4.9	605	1 UBP2 HUMAN	O75604 homo sapien
15	412	4.8	353	1 UBP2 MOUSE	O88623 mus musculus
16	371	4.3	926	1 UBP4 YEAST	P32571 saccharomyc
17	343.5	4.0	913	1 UBP4 HUMAN	Q92k6 homo sapien
18	330	3.9	805	1 UBP5 YEAST	P39944 saccharomyc
19	323.5	3.8	566	1 UBP1 MOUSE	Q9qz16 mus musculus
20	321	3.7	565	1 UBP1 HUMAN	Q9uk80 homo sapien
21	303	3.5	438	1 UBP4 SCHPO	O60139 schizosacch
22	294	3.4	823	1 UBP6 HUMAN	Q95t5 homo sapien
23	285	3.3	520	1 UBP3 MOUSE	Q81w36 mus musculus
24	271	3.2	593	1 UBP6 HUMAN	Q8upt9 homo sapien
25	259	3.0	445	1 UBP6 CANFA	Q01988 canis fami
26	249.5	2.9	1071	1 UBP7 YEAST	P40453 saccharomyc
27	245.5	2.9	521	1 UBP3 HUMAN	Q9y614 homo sapien
28	211.5	2.5	717	1 UBP3 YEAST	P36026 saccharomyc
29	210.5	2.5	913	1 UBP6 HUMAN	Q85xu7 homo sapien
30	206	2.4	898	1 UBP6 DROME	Q24574 drosophila
31	203.5	2.4	922	1 UBP7 HUMAN	Q9hbj7 homo sapien
32	200	2.3	526	1 UBP6 MOUSE	Q61068 mus musculus
33	190	2.2	754	1 UBP9 YEAST	P39967 saccharomyc

RESULT 1

ID	UBP6 HUMAN	STANDARD;	PRT;	1089 AA.
AC	P35125;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	Ubiquitin carboxyl-terminal hydrolase 6 (EC 3.1.2.15) (Ubiquitin thioesterase 6) (Ubiquitin-specific processing protease 6)			
DE	(deubiquitinating enzyme 6) (Proto-oncogene TRE-2).			
GN	USP6 OR TRE2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92228503; PubMed=1565468;			
RA	Nakamura T., Hillova J., Mariage-Samson R., Onno M., Huebner K., Cannizzaro L.A., Boghosian-Seil L., Croce C.M., Hill M.;			
RA	"A novel transcriptional unit of the tre oncogene widely expressed in human cancer cells."			
RT	human cancer cells."			
RL	Oncogene 7:733-741(1992).			
RN	[2]			
RP	CHARACTERIZATION.			
RP	MEDLINE=94067315; PubMed=8247125;			
RX	Papa F.R., Hochstrasser M.;			
RA	"The yeast DOA4 gene encodes a deubiquitinating enzyme related to a product of the human tre-2 oncogene."			
RT	Nature 366:313-319(1993).			
RL	Nature 366:313-319(1993).			
CC	- FUNCTION: HAS AN ATP-INDEPENDENT ISOPEPTIDASE ACTIVITY, CLEAVING AT THE CARBOXYL TERMINUS OF THE UBIQUITIN MOIETY.			
CC	- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O = ubiquitin + a thiol.			
CC	- DISEASE: INVOLVED IN TUMORIGENESIS.			
CC	- SIMILARITY: Belongs to peptidase family C19.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X63547; CAA45111.1; ..			
DR	EMBL; X63546; CAA45109.1; ALT_INIT.			
DR	PIR; S57868; S22158.			
DR	MEROPS; C19.009; ..			
DR	Genew; HGNC:12629; USP6.			
DR	MIM; 604334; ..			
DR	GO; GO:0004212; P-lysosomal cysteine-type endopeptidase; TAS.			
DR	GO; GO:0003676; Nucleic acid binding activity; TAS.			
DR	GO; GO:0007048; P-oncogenesis; TAS.			
DR	InterPro; IPR001394; UCH-2.			
DR	Pfam; PF00443; UCH; 1.			

Q9es63 mus musculus
Q3p785 schizosacch
Q94782 homo sapien
Q91614 xenopus lae
P36610 homo sapien
Q99mx1 mus musculus
O75317 homo sapien
P42326 drosophila
Q09711 schizosacch
Q96ru2 homo sapien
P38237 saccharomyc
P79880 gallus gall

34 186 2.2 869 1 UBPT MOUSE
35 184.5 2.2 875 1 UBPT SCHPO
36 183 2.1 785 1 UBPT HUMAN
37 173.5 2.0 189 1 NCS1_XENLA
38 172.5 2.0 189 1 NCS1 HUMAN
39 172.5 2.0 835 1 UBPO MOUSE
40 169 2.0 355 1 UBPC HUMAN
41 167 2.0 189 1 NCAH DROME
42 167 2.0 189 1 NCS1 SCHPO
43 166 1.9 1077 1 UBPE HUMAN
44 165 1.9 803 1 UBPE YEAST
45 164 1.9 198 1 GCAI_CHICK

ALIGNMENTS

DR PROSITE; PS00972; UCH 2.1; 1.
DR PROSITE; PS00973; UCH 2.2; 1.
DR PROSITE; PS02335; UCH 2.3; 1.
KW Proto-oncogene; Ubl conjugation pathway; Hydrolase; Thiol protease;
KW Multigene family.
FT ACT_SITE 438 BY SIMILARITY.
FT ACT_SITE 1003 BY SIMILARITY.
FT ACT_SITE 1011 BY SIMILARITY.
SQ SEQUENCE 1089 AA; 121883 MW; 6228BCFBED5B980 CRC64;
Query Match 52.9%; Score 4531; DB 1; Length 1089;
Best Local Similarity 93.7%; Pred. No. 1.1e-278;
Matches 851; Conservative 22; Mismatches 31; Indels 4; Gaps 1;
QY 701 EVNKDMSPEEMSFANSKIDRHVPTKEKATGLNLTGNTCFWNSISQCVNTPQTQ 760
DB 182 EVNKDMSPEEMSFANSKIDRQVPTKEKATGLNLTGNTCFWNSISQCVNTPQTQ 241
QY 761 YFISGRLYELARTNPIDGKGMKACYGDLVQELMSGTQKNVAPLKLRTWIAKYAPRNG 820
DB 242 YFISGRLYELARTNPIDGKGMKACYGDLVQELMSGTQKNVAPLKLRTWIAKYAPRNG 301
QY 821 FQODSQELLAFILDLGLHEDLNVRVHEKPYVELKSDGRPDWEVAEADNHLRRNRSTV 880
DB 302 FQODSQELLAFILDLGLHEDLNVRVHEKPYVELKSDGRPDWEVAEADNHLRRNRSTV 361
QY 881 DLPHGLRQVQKCTCGHISVRFDPNPLSLPLMDSYMHLITVTKLDGTPVRYGLRL 940
DB 362 DLPHGLRQVQKCTCGHISVRFDPNPLSLPLMDSYMHLITVTKLDGTPVRYGLRL 421
QY 941 NMDEKVTGLKQSLDLGLNLSQIILAEVHSGNINKNFQDNQKVLVSGFLCAFEI 1000
DB 422 NMDEKVTGLKQSLDLGLNLSQIILAEVHSGNINKNFQDNQKVLVSGFLCAFEI 481
QY 1001 VPSIASSTQTDFFSSPSTNEMFTLTNGDLPRIPIFNGMPTNVPVPGTEKFTNGMV 1060
DB 482 VPSIASSTQTDFFSSPSTNEMFTLTNGDLPRIPIFNGMPTNVPVPGTEKFTNGMV 541
QY 1061 NGHMPSLDPTGVIIVHRRKQWTELYFLSSQNRSLFGMLPLVICTVHTKQKLYD 1120
DB 542 NGHMPSLDPTGVIIVHRRKQWTELYFLSSQNRSLFGMLPLVICTVHTKQKLYD 601
QY 1121 AVMIQVSLASLPPEASNAHQDSDSGYQPTFLRVQKDGNSCAWCFWFCRCK 1180
DB 602 AVMIQVSLASLPPEASNAHQDSDSGYQPTFLRVQKDGNSCAWCFWFCRCK 661
QY 1181 IDCGRDRAFINAYTAVDWDPTALHLRYQTQSERVVDHESVQSRRAQAPINLDSCLR 1240
DB 662 IDCGRDRAFINAYTAVDWDPTALHLRYQTQSERVVDHESVQSRRAQAPINLDSCLR 721
QY 1241 AFTSEELGENEMYCSKCKTHCLATKLDLWRLPPLIILHKKRQFVNGEWIKSQIVK 1300
DB 722 AFTSEELGENEMYCSKCKTHCLATKLDLWRLPPLIILHKKRQFVNDQWIKSQIVR 781
QY 1301 FRESFDSAFVLPDPALCQHKPTTQGBELSEPRILAREVKKYVDAQSSAGEEDVLLSK 1360
DB 782 FRESFDSAFVLPDPALCQHKPTTQGBELSEPRILAREVKKYVDAQSSAGEEDVLLSK 841
QY 1361 SPSSLISANISSPKGSPSSRSKSGTSCPSKNSPNSPRTILGSKGLRLPQIGSKVKL 1420
DB 842 SPSSLISANISSPKGSPSSRSKSGTSCPSKNSPNSPRTILGSKGLRLPQIGSKVKP 901
QY 1421 SSSKNLDAKENGAGQICELADALSRGHVGGSGQPELVTPQDHEVALANGFLYEACG 1480
DB 902 SSSKNLDAKENGAGQICELADALSRGHVGGSGQPELVTPQDHEVALANGFLYEACG 961
QY 1481 N---GYSGNGLGNHSESDTDQDRETRIKPIYLNLYAISCHSILGCHGVTVYAKNPNC 1536
DB 962 NCGGNGNGLGNHSESDTDQDRETRIKPIYLNLYAISCHSILGCHGVTVYAKNPNC 1021
QY 1537 KWCYNDSSCKELHPDEITDSAYILFYEQQIDYVAQPLPKTDGKWADTSSDDESD 1596
DB 1022 KWCYNDSSCKELHPDEITDSAYILFYEQQIDYVAQPLPKTDGKWADTSSDDESD 1081

QY 1597 YKCYVQLQ 1604
DB 1082 YEKYSMLQ 1089
RESULT 2
UBPF_MOUSE
ID_UBPF_MOUSE STANDARD; PRT; 991 AA.
AC Q8R5H1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 15 (EC 3.1.2.15) (Ubiquitin
thiolesterase 15) (Ubiquitin-specific processing protease 15)
DE (Deubiquitinating enzyme 15).
GN USP15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SQ SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND TISSUE SPECIFICITY.
RC STRAIN=C57BL/6;
RX MEDLINE=22419901; PubMed=12532266;
RA Angelats C., Wang X.-W., Jermin L.S., Copeland N.G., Jenkins N.A.,
RA Baker R.T.;
RT "Isolation and characterization of the mouse ubiquitin-specific
protease Usp15".
RL Mamm. Genome 14:31-46 (2003).
RN [2]
SQ SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina, Testis, and Thymus;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Akawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Sado K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Sava T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
ubiquitin + a thiol.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Experimental confirmation may be lacking for some
isoforms:
CC Name=1;
CC IsoId=Q8R5H1-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8R5H1-2; Sequence=VSP_005262, VSP_005263;
CC Name=3;
CC IsoId=Q8R5H1-3; Sequence=VSP_005264;
CC Name=4;
CC IsoId=Q8R5H1-4; Sequence=VSP_005265, VSP_005266;
CC -1- TISSUE SPECIFICITY: Widely expressed with highest levels in
testis, heart and liver.
CC -1- SIMILARITY: Belongs to peptidase family C19.

This SWISS-PROT entry is copyright. It is produced through a collaboration

Db 799 CPNCKGQOATKFDLWLSPLKILVHLKFSY-NRYWRDKLTVVEFFIRGLNMFSEF--- 854
 QY 1315 RDPALCOHKLPTPGQDELSEPRILAREVKVDAQSSAGEEDVLLSKSPSSIANIISPK 1374
 Db 855 ----- 854
 QY 1375 GSPSSSRKSGTSCPSKSNSSPNSPRTLGRSKRLRLPQIGSKNKLSSKENLDASKENG 1434
 Db 855 ----- 854
 QY 1435 AQOICELADALSRGHVLSGQSELPVTPQDHEVALANGFLYEHAACNGYNGQLGNHSEE 1494
 Db 855 -----VCNLS----- 859
 QY 1495 DSTDDQREDTRIKP-IYNLYALSCHSGILGGHYVYAKNP-NCKWYCVNDSSCKELHPD 1552
 Db 860 -----ARPYVDLIAVSNHYGANGVGHYAYAKNKLNGWYFYDSDNVLASED 908
 QY 1553 EIDTDSAYILFYEQGIDYAOFLPKTDGKQVAD-----TSSMDEDFSD 1596
 Db 909 QIVTKAAVYLFYQRDRDFYK-TPSLSSSGSDGGRPSQSGFGDD 955

RESULT 4
 UBP4_MOUSE
 ID UBP4_MOUSE STANDARD; PRT; 962 AA.
 AC P35123; OS4704;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 4 (EC 3.1.2.15) (Ubiquitin
 thiolesterase 4) Ubiquitin-specific processing protease 4)
 DE (Deubiquitinating enzyme 4) (Ubiquitous nuclear protein).
 GN USP4 OR UNP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1] SEQUENCE FROM N.A.
 RP TISSUE=Testis;
 RC MEDLINE=98267137; PubMed=9602026;
 RA di Fruscio M., Gilchrist C.A., Baker R.T., Gray D.A.;
 RT "Genomic structure of Unp, a murine gene encoding a ubiquitin-specific
 protease";
 RL Biochim. Biophys. Acta 1398:9-17(1998).
 RN [2] SEQUENCE FROM N.A.
 RP STRAIN=129;
 RC Di Fruscio M., Gilchrist C.A., Baker R.T., Gray D.A.;
 RA "Genomic structure of Unp, a murine gene encoding a ubiquitin-specific
 protease";
 RT Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RL [3] PRELIMINARY SEQUENCE OF 169-962 FROM N.A.
 RN MEDLINE=93330575; PubMed=8336951;
 RX Gupta K., Copeland N.G., Gilbert D.J., Jenkins N.A., Gray D.A.;
 RA "Unp, a mouse gene related to the tre oncogene";
 RT Oncogene 8:2307-2310(1993).
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 ubiquitin + a thiol.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- DEVELOPMENTAL STAGE: OVEREXPRESSION LEADS TO ONCOGENIC
 CC TRANSFORMATION OF NIH 3T3 CELLS
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; L00681; AAB82339.1; --
 DR EMBL; AF026469; AAC53587.1; --
 DR MEROPS; C19.010; --
 DR MGI; MGI:98905; Usp4.
 DR InterPro; IPR006615; DUSP.
 DR InterPro; IPR001394; UCH-2.
 DR Pfam; PF00443; UCH.1.
 DR SMART; SM00895; DUSP; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family;
 KW Proto-oncogene; Nuclear protein.
 FT ACT_SITE 311 311 BY SIMILARITY.
 FT ACT_SITE 872 872 BY SIMILARITY.
 FT ACT_SITE 880 880 BY SIMILARITY.
 FT DOMAIN 765 772 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT CONFLICT 292 292 A -> P (IN REF. 2).
 SQ SEQUENCE 962 AA; 108280 MW; 9553CAD9B4857360 CRC64;
 Query Match 12.7%; Score 1086; DB 1; Length 962;
 Best Local Similarity 24.2%; Pred. No. 9.8e-61;
 Matches 322; Conservative 164; Mismatches 322; Indels 522; Gaps 38;
 QY 390 LQAGHNWFIISMOWQWQWKEYKYKIDANPVVIBPSSVLNGKYSFGTAHPMQVEDRIGS 449
 Db 28 LQGAQWYLDSRNFQWKKYVGF-----SWDMYNG----- 60
 QY 450 SLSYVNTTEKFSEDNISTASEASETAGSGFLYSATGADVCFARQHTSDNNNQCILLGAN 509
 Db 61 -----EHNLF----- 65
 QY 510 GNILLHLNPKGAINQPLVTOEPVKATSLTEGRLKRTPLQHGRDYEVVPEVWEA 569
 Db 66 -----POPIDNSGLFS-DP-----ESOTLKE--HLIDELDYLVPAEAMNK 103
 QY 570 LYHWYGANALPRPVKSKSDIPELELPPR-----YLLFLRQOPATRTQGSNIWVMNG 623
 Db 104 LLNMYGC-VEGQOPIVRKVVD-----GLFVKHCKVEVILLEKL----- 142
 QY 624 NVPSNPAPLKRVLAYTGC-FSRMQIKETHEVLSORLRI-KEEDRLWLYNSENYLTLD 681
 Db 143 ---CENSDDTNVLS---CHFSKADTIATEKEMRKLFPNPAERETRLWNKYSNTYEQLS 196
 QY 682 DEHKLEYLKIODEQHLVIEVRNKMSPPE----- 712
 Db 197 KLDNTIQDAGLYOGVIVIEPQNEEDGTWPQSLQSKSTAPSRNFTTSKPSASPYCVS 256
 QY 713 MGFIAN-----SSKIDR-----HKVPT---EKATGLSNLGNTCFMNSS 748
 Db 257 ASLIANGDSTNSGMSHSGVSRGSGFSGFASYNQCQAPSPHIQPGCLGLNLGNTCFMNSA 316
 QY 749 IOCVSNTOPLTOYFISGRHLYELNRTNPIGMKGMKACVGDIVQELWSGTQKNVAPLKL 808
 Db 317 LQCLSNAPLTFYFLKDEYAEINRDNPLGMKGEAEAYAEILKOMWGRDTHVAPRMFK 376
 QY 809 WTIAKYAPRNFQFOODSQELLAFLLDGLHEDLNVRHEKPYVELKSDGRPDWEVAEAW 868
 Db 377 TQVGRFAPQFSGYQOQSQELLAFLLDGLHEDLNVRKPKVLEPKDANGRPDAVVAKEAW 436
 QY 869 DNEHLRNSIVVDLPHGQLRSQVKCTCGHISVRDPPNPLSLPLPMDSYMHEITVLK 928
 Db 437 ENHRLRNSVIVDTFHLGFLFKSLVQPECAKSVTFDPFCYLTLPPLKDKRIMEVFLVPA 496
 QY 929 D-GTTPVRYGLRNLNDEKTYGLKKLSDLCLGNSQIILAEVHGSNI-KNFPQDNQKRL 986
 Db 497 DQCRPIQRYVTPLMGALSDICEALSLSGLAAENMVVTDVYHNRPHKIFQMDGLSHI 556
 QY 987 SVSGFLCAFEIPVPVSPISASSPTQDFSSSPSTNEMETLTNGDLPRPIFGMPNTV 1046
 Db 557 TPRDDIFVYEV-----CNTSDMGSECITL-----PVYF----- 584

RA Whitehead S.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBSJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stables M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL; U44839; AAC50450.1; -;
 DR EMBL; AL096791; CAD20056.1; -;
 DR EMBL; BC000350; AAH00350.1; -;
 DR MEROPS; C19.014; -;
 DR Genew; HGNC:12609; USP11.
 DR MIM; 300050; -;
 DR GO; GO:0004197; P:cysteine-type endopeptidase activity; TAS.
 DR GO; GO:0004843; P:ubiquitin-specific protease activity; TAS.
 DR Interpro; IPR001394; UCH-2.
 DR Pfam; PF00443; UCH; 1.
 DR PROSITE; PS00972; UCH 2_1; 1.
 DR PROSITE; PS00973; UCH 2_2; 1.
 DR PROSITE; PS0235; UCH 2_3; 1.
 DR Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT_SITE 53 53 BY SIMILARITY.
 FT ACT_SITE 607 607 BY SIMILARITY.
 FT ACT_SITE 615 615 BY SIMILARITY.
 FT CONFLICT 216 216 P -> L (IN REF. 3).
 SQ SEQUENCE 690 AA; 79128 MW; A7DDE9BD3ADE259 CRC64;
 Query Match 10.6%; Score 903.5; DB 1; Length 690;
 Best Local Similarity 25.9%; Pred. No. 2,1e-49;
 Matches 249; Conservative 122; Mismatches 231; Indels 359; Gaps 23;
 QY 700 IEVRNKMDSWPEMSFIANSKIDRHK-VPTKKGATGLSNLGNCTCFMNSSIQCVSNTPQL 758
 DB 1 METRKKDGTWPAQLHVNWNNSDEEDFKQPGICGLTGLNCTCFMNSALQCLSNVPL 60
 QY 759 TQYFTSGRHLVELNTNFGMKGHAKYGDVLQVELWSGTQKNVAPLKLRTWIAKAPRF 818
 DB 61 TSYFLNCCLEELNFRNPLMGKEIAEAYADLVKQAWSGHSRISVPHVFNKVGHFASQF 120
 QY 819 NGFQOQDSOELLAFLLDGLHDLNVRHVKPYVELKDSGRPDWEAAEADNHLRRNRSI 878
 DB 121 LQYQHDSDSLLSFLLDGLHDLNVRHVKPYVELKDSGRPDWEAAEADNHLRRNRSV 180

QY 879 VVDLHFGQLRSQVKCKTCGHSVRDPENFLSLPLPMDSYMHELEITVTKLD-GTTPVRVG 937
 DB 181 IVDTHGLFRSLVCPDCGNVSVTFDFPCYLSVPLPSHKRVLEVFPPMDPRRPEQHR 240
 QY 938 LRINMDEKTYGLKKQLSDCLGNLSQILALAEYHGSNIKFNPDQDNQKVLVSGLCAPEI 997
 DB 241 LVVPKKGKISDLCAVALSKHTGISPERMMVADVFSHR-----FYKLYQL 283
 QY 998 PVPVSISASSPTQDTDFSSSPSTNEMFILTNG-----DLPRPIFINGMPTVV 1047
 DB 284 EEPPLSIL-----DRDIFVYVSGRIEAGSREDIVVPVYLRERP----- 326
 QY 1048 PCCTEKNTNG-----MVNGH--MPSLPDPSPTGYIIAVERKMMRTELYFLSSQKRPRLF 1101
 DB 327 ----ARDYNSYVGLMFLGHLPLVSPDRFTW-----SNPE----- 355
 QY 1102 GMPLIVPCTVTRKLDLYDAVVIQVSR-----EGLNVLMYRLSRYVTKNSDDEDDGDEKDEEDKDPVGPSTG 1128
 DB 356 -----EGLNVLMYRLSRYVTKNSDDEDDGDEKDEEDKDPVGPSTG 401
 QY 1129 -LASPLPQO--EASNHAQDC---DMSG-YOYP-----FTLRVYQKGNCSAWCPW 1172
 DB 402 SLRDEPEQAGPSSGVTRPCFLDNCIGTSQNPVRRRKQLFTLQTVNSNGTS----- 455
 QY 1173 YRFRCKIDCGEDRA-----FIGNAYIAVDWDPALHLRYQTSQERVVDEHESVEQSR 1227
 DB 456 -----DRTSPSEVHAQPYIAIDWEPKMKRYVDEVEAGYVHKDCVGYMK 502
 QY 1228 AQAEPIINDSLRAFTSEELGENEMYCKCKTHCLATKLDLWLPPLIILHLKRFQF 1287
 DB 503 --KAPVRLQECLELFTVTLEKENFTWYCPSCQKQLATKLDLWLPPLIILHLKRF 560
 QY 1288 VNGRWTKSKIVKFPRESFDPASFLVPRDPALCOHKPLTPQGDSELPRLAREVKYKDA 1347
 DB 561 TKFRSEKLDLVEFPTRDLDFSEFVI-----QPQNE----- 591
 QY 1348 QSSAGEEDVLLSKSPSSLSANIISPKSPSSRSKSGTSCPSCKNSPNSPRTILGRSG 1407
 DB 592 -----SNPE----- 595
 QY 1408 RLRLPQIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLSQSPQLVTPQDHEVA 1467
 DB 592 -----SNPE----- 595
 QY 1468 LANGFLYEHAACGVNSQNGQLGNHSEEDSTDQREDTRIKPIYNDLYAISHGILGGHY 1527
 DB 596 -----LYK-----YDLIAVSNHYGMRDGHY 616
 QY 1528 VTYAKNPNC-KWVCYNDSSCKELHPDEITDTSAYILFYEQQIGIDYAQFLPKTDGKKMADT 1586
 DB 617 TTFACNKGSGWHYFDDNSVSPVNEQIESKAAVYLFYQRO--DVARRLLSPAGSSGAPA 674
 QY 1587 S 1587
 DB 675 S 675
 RESULT 7
 UBPE_MOUSE
 ID UBPE_MOUSE STANDARD; PRT; 699 AA.
 AC Q99K46;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 11 (EC 3.1.2.15) (Ubiquitin
 DE thiolesterase 11) (Ubiquitin-specific processing protease 11)
 DE (Deubiquitinating enzyme 11).
 GN USP11.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;

RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RA "The genome sequence of Schizosaccharomyces pombe.";
RT Nature 415:871-880(2002).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AL023776; CAAL19303.1; -;
CC PIR; T41006; T41006
CC GeneDB Spombe; SPCC1494.05c; -;
CC InterPro; IPR006615; DUSP.
CC InterPro; IPR001394; UCH-2.
CC Pfam; PF00443; UCH; 1.
CC SMART; SM00695; DUSP; 1.
CC PROSITE; PS00972; UCH_2_1; 1.
CC PROSITE; PS00973; UCH_2_2; 1.
CC PROSITE; PS02035; UCH_2_3; 1.
CC Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
KW ACT SITE 320
FT BY SIMILARITY.
FT ACT SITE 927
FT BY SIMILARITY.
FT ACT SITE 935
FT BY SIMILARITY.
SQ SEQUENCE 979 AA; 111967 MW; A705444BCDB437D2 CRC64;

Query Match 9.2%; Score 787.5; DB 1; Length 979;
Best Local Similarity 22.9%; Pred. No. 7.8e-42;
Matches 283; Conservative 162; Mismatches 414; Indels 377; Gaps 37;

QY 432 SFGTAAPMEQVEDRIGSSISVYNTTEKFSNDISTASEATAGSGFLYSATPGADVCF 491
Db 19 SLSESSQSSNNWDIISQKSIISLGDAE--ISKNLPSIAKQKQLIG----- 61
QY 492 ARQNTSDNNQCLLGANGNILLHLNPK-----PGAINDQPLVTOEPVKATS 539
Db 62 ----ELVNNQPELQGVNDVYLSYVYERLCSYLAEDGPPGVPVQEDIA----- 108
QY 540 LTLGGRLKRTFQLHGRDVEVPEVYRALYHWYG--ANLALRPVTKNSKTDIPEL--E 596
Db 109 -DLTGILK--PDLOEIDFIIRSDVWLLVRYGLKGPFFRETNLGSESHPLHVE 165
QY 597 LFPRLYLFLRQOPATRTQSQNIWNNMGNVSPNAPLKRVLAYTGCFSRMOTIKEIHEYL 655
Db 166 VYPPIFSLTLSLTNAVDANES-----HKPKKISL-----SSKSTLEDLEGV 207
QY 656 SQRRLRIKEDNRMLYNSENVLTLDDEHKLKYLKIQDQOHLVIEVRNKM----- 707
Db 208 KYTSLSPSDQRLRWVPTDQPLHRTIDPS---SPKINSKEIIDFLEKSKTLVELGMDSS 264

QY 708 -----SWPEE-----MSPI-----ANSSKIDRHVPTKCATGLSNLNTCF 744
Db 265 CSLVAECMINETFPVDRALRLQFLIQORNNSSBEQKQKRVF---GTGSLNLGNTCY 321
QY 745 MNSSIQCVNTQPTQYFISGRHLYELNRTPIGWKGMKACYGDLVQELWSGTQKNVAP 804
Db 322 MNSALQCLTHRELDFTTSDEWKNQVNESPLNGWGQVAFISLILKSLISPEHSFAP 381
QY 805 LKLRTIAYAPRNGFQQQSQELLAPLLDGLHEDLNRVHEKPYV---ELKDSGDGRDW 861
Db 382 RQKATIGKFNHSLFGYQQDSQSFLAPLLDGLHEDLNRIYQKPYTKPDIYEVEDEEK 441
QY 862 EYAAEAWNHLRRNRSIVVDLFGQLRSOVKCKTCGHSVRFDPDFNFIPLPLPWSYML 921
Db 442 NTAECWRLKLRNDSLIVDLFGQYRSITVPCVNTVSTITDFDFMDLTLPPLPKQVWSH 501
QY 922 EITVKLD-GTTPVRYGRLNMD-----KYTLKKQLSDLCGLNSEQILLAEVHGSN 973
Db 502 TVTFIPADTNTLPLAIEVVLESKAATIEDLVKYVAEKSGCSDY-----RKILVTETYKGR 556
QY 974 IKNF-PODNQKVRLSVG-----FLCAFEIP-----VPSVISASSPTQTD-FSSSP 1018
Db 557 FYRFTQLSKSLMEISEDEIYLYELRPYEDGSDILVPVYHISDDSTNSANSYMSR 616
QY 1019 STNEMFTL-TTNGDLPRPIFIPNGMPTNVVPCGTEKQFTN-----GMVNGHMPSLPDSPP 1072
Db 617 DFHGFPVLQSDNEVTDASFISEKLKLYQQFTTLKNLKNIDSLSESLGH-----EDEQV 672
QY 1073 TGYIIVARKMRTELTYLLSSQNR-----PSLFGMPLI-VPCVTHTKKDLYDAWIVQS 1127
Db 673 QKGPLDVMDSHQSOTPLFEMRVFHDRFEKIPITGWNMSVSNLPLTERDKD----- 722
QY 1128 RLASPLPQEAENHAQDCDDSMGYQYPTLRVVOGDGNSCAWCPWYRCRCCKIDCGS-D 1186
Db 723 -LESTVDPDLAHSIEEDDSEFKDVAFGSYEPKSNEN-----TKLTAKEND 769
QY 1187 RAFIGNAYIADVDPALHLRYQT-----SQERVVDEHESVEQSRRAQAP 1232
Db 770 RLLIQGLLVCEWPEKSYQFVSAPSSPQMSRSLWLESKITLSDKKDSEDSR-----T 824
QY 1233 INLDSCLRAFTSEELGENEYCYCKKTHCLATKLDLWLPPLIILHLARFQVNGRW 1292
Db 825 IYLNCLDFEFKTEQLGESEDPWYCTCKEPRQASQOMEIMRCPELILFHLKRFSSERRFR 884
QY 1293 IKSQRIKVFPPRESFDPASFLVRDPALCOHKPLTPQGDSELPRLAREVKVDAQSSAG 1352
Db 885 DKIDDLVEFP----- 894
QY 1353 EDVLLSKSPSLSANIISSPKGSPSSRSKSGTSCPSKNSPNSPRTLGRSGRLRLP 1412
Db 895 ----- 894
QY 1413 QIGSKNKLSSKENLDASKENGAGQICELADALSRGHVGGSPQLVTPQDHEVALANGF 1472
Db 895 -----IDNLDMSRTGSYKLS----- 911
QY 1473 LYEHACNGVYNGOLGNHSEBSTDQDRETRIKPIYNYLVAISCHSGILGGHYVYAK 1532
Db 912 -----KENPKL--IYELYAVDNHYGGLGGHYTAFK 941
QY 1533 NP-NCKWYCYNDSCKELHPDIDTDSAYILFYEQO 1567
Db 942 NPDNGQFICFDDSRVTPVCPETWTSAAYLLFYRRK 977

RESULT 9

UBPC YEAST
ID UBPC YEAST STANDARD; PRT; 1254 AA.
AC P39538;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Ubiquitin carboxyl-terminal hydrolase 12 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 12) (Ubiquitin-specific processing protease 12)
DE (Deubiquitinating enzyme 12).
GN UBPI2 OR YUL197W OR J0340.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
EX MEDLINE=95274326; PubMed=7754713;
RA Fournelle B., Cosser F., Goffeau A.;
RT "The sequence of a 36 kb segment on the left arm of yeast chromosome
RT x identifies 24 open reading frames including NUC1, PRP21 (Spp91),
RT CDC6, CRY2, the gene for S24, a homologue to the aconitase gene ACO1
RT and two homologues to chromosome III genes.";
RL Yeast 10:1235-1249(1994).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X77688; CAA54762.1; -
CC DR EMBL: Z49472; CAA89492.1; -
CC DR PIR: S46636; S46636.
CC DR MEROPS: C19.0PW; -
CC DR SGD: S0003733; UBF12.
CC DR InterPro: IPR006615; DUSP.
CC DR InterPro: IPR001394; UCH-2.
CC DR Pfam: PF00443; UCH; 1.
CC DR SMART: SM00695; DUSP; 1.
CC DR PROSITE: PS00972; UCH_2_1; 1.
CC DR PROSITE: PS00973; UCH_2_2; 1.
CC DR PROSITE: PS0235; UCH_2_3; 1.
CC KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT ACT_SITE 373 373 BY SIMILARITY.
FT ACT_SITE 1060 1060 BY SIMILARITY.
FT ACT_SITE 1068 1068 BY SIMILARITY.
SQ SEQUENCE 1254 AA; 143191 MW; 349AFA4C4CEE0EAS CRC64;

Query Match 8.6%; Score 738.5; DB 1; Length 1254;
Best Local Similarity 21.8%; Pred. No. 1.5e-38;
Matches 296; Conservative 159; Mismatches 384; Indels 521; Gaps 47;

QY 445 DRIGSLSYVNTTEKFSNDISTAS-----EASSETAGSLYSATPGADVCFARQHTS 498
DB 29 DELGVDL--MNVLDK--DEIKQESVPVSDREIEDTESDASAVSFASANELIAEPHAAS 84
QY 499 DNNNQCLLGANG----NILLHLNPKCAIDNPLVTOEPVKATSLT-LEGRLKRTPOL 553
DB 85 ETN-----LGTNGQGRNVL-----EQORDVVARLIEENKETQKGGDKVCIYPKV 129
QY 554 IHG-----RDYE-----MVPEPYWRALYHWYGANL 578
DB 130 WYDKFPDPVTDPEDIGPINTRMICRDPENFVLEDYNRCPVLSIAEPVFNFLSIYGMTS 189
QY 579 ALPRVINKSTDIPELIFPRY-----LLFLRQOPATRCQ-----SNTWNGNVPS 627
DB 190 G-SYVPTNLVNTQTGTETEYNKWFRLHYLTKQDGRKRREGQDDSIYLSM----- 243
QY 628 ENAPLKRVLATGCFSRMOTIKEIHE-----YLSQRLIKEEDMLW-----LYNS 673
DB 244 -----SALNLRVDLVEKSMNLFPEKADHLVNAVDFKLIWFSEGSDIATD 288
QY 674 ENYLTLLDD--EDHKLEVLKI-----ODEQHLVIEVR--NKDMSWP 710

DB 289 SNVSTELNSSYEITPLOFLELPKILLIPDMFENRLDKITSPSDLVIEIKPIEGNHWP 348
QY 711 EEMSFIANSSKIDRHVKPTEKATGATLSNLTGNTCFMNSSIQCVSNTOPLTQYFISGRHLVE 770
DB 349 -----SNYFAYNKLBPASGTTGLVNLGNTCYMNSALQCLVHPQLRDYFLDYDGEDE 400
QY 771 LNETNPIMKGMKACVGLVQELW-----SGTQKVA--PLKLRWTIAKYAVRENGFQQQ 824
DB 401 INEENPLGTHGVARAFSDLVQKLFQNRMSIQNRNAAPPSMFKSTIGHFNSMFSYMQQ 460
QY 825 DSQELLAFLDLGLHEDLNVRHEKPYVE---LKDSDGRPDWEV---AAEAWNDHLRNRNS 877
DB 461 DSQELFLAFLDLGLHEDLNRIIKEYTEKPSLSPGDDVNDVNVVKLADDTWEMHLKRNS 520
QY 878 IYVDLPHGQLRSQVKCKTCHISVRPDPENFLSLPLPMDSYHLEITVILKOGTTPVRVG 937
DB 521 VITDLFVGMKSTLYCPEQCNVITFDYNDVTLPLVDVTVDWTKIKIPPMN-SPILLLE 579
QY 938 LRLNMDKTYTGLKKQLSDLCGLNSEQIILLAEVHGNIKNFQDNOKVRLSVSGFLCAFEI 997
DB 580 VELSKSSTYMDLKNVYVGMKSLDPTLFCCEIFSNQI-----YVNYESTESNAQFLTQEL 635
QY 998 PVPVSPISASSPTQDTDFSSSPSTNE-----MFTLTNGDL 1032
DB 636 IKPADDAVI-----FYELPVTNDNEVIVPVLNTRIEKGYKNAMLFGVPFFITLKEDEL 687
QY 1033 PRIFITPQMGPTVW-----PCCTEKNTFTNGVMVNGHMPSI-----PDSPTFGY-- 1075
DB 688 NNPGATRMKLNRFVHLSGGYIPFPEPVGNRTDFADA-----FPLLVEKYPDVEFEQYKD 742
QY 1076 -----IIAHRKMMRTLYFLSSQKNRPSLFGMPLI----- 1106
DB 743 ILOYTSIKVTDKDKSPFSIKILSVK-----EQQFASNNRTGFN-FWTFISQLNLDKATD 796
QY 1107 VPCVHTTRKKDLY-----DAYWIOV-----SLASP-----LP 1134
DB 797 IDKLEDDVVKDIYVSSSLVDCAEGLMVDDDEGDTGSEAKNFSPFOSGDDDEENKETVT 856
QY 1135 PQEASNAHQDCDDSMGY-----TDDVEDASTEPELTDKPEALDKIKDSLTSTPFAISM--N 912
DB 857 NNENVTNDRDEDMELTDDVEDASTEPELTDKPEALDKIKDSLTSTPFAISM--N 912
QY 1166 SCWCWYFRCGKIDCGEDRAFTGNVAVDWPATLHLRYQTSQERVVDDEHSVBOS 1225
DB 913 DIIVCEW-----SELGSNEAFSDDK--IYNWENPA-----TLPNKELE-ENAKLERS 955
QY 1226 RRAQAEPINLDSCLRAFTSEBELGENEMYCCKTHCLATKLDLWRLPPTLIILHLKRF 1285
DB 956 -NAKERTIITLDDCLQLFSKPEILGLTDSWYCTCKEHRQATKQIQIWNTPDILLHLKRF 1014
QY 1286 QVNGRWIKSQKIVKPPRESFDSAPLVRDPALCQHKPLTPQGDDELSEPRILAREVKV 1345
DB 1015 ESQSPSPKIDATVNFPIITDLSRYVVKD----- 1045
QY 1346 DAQSSAGEEDVLLSKSPSSLSANISSPKGSSSRKSGTSCFSSKNKSNPSSPRTLGHS 1405
DB 1046 -----DPRGL----- 1050
QY 1406 KGLRLPQIGSKNKLSSKENLDAKENGAGQICELADALSRGHVGGSQPELVTPQDHE 1465
DB 1051 ----- 1050
QY 1466 VALANGFLYEACNGYNGQIKNHSEBDDTDRETRIKPTIYNLYAISCHSILGGG 1525
DB 1051 -----IYDLAYVDNHYGGLGGG 1067
QY 1526 HVTYAKN--PNCKWVCYNDSSCKELHPEIDTDSAVILFY 1564
DB 1068 HTAYVKNFADKNYTFDDSRVTETAPENSIAGSAYILFY 1107

RESULT 10

UBP1_SCHPO STANDARD; PRT; 849 AA.

AC Q9USM5, 8.1%; Score 693.5; DB 1; Length 849;
Best Local Similarity 21.3%; Pred. No. 5.6e-36;
Matches 233; Conservative 160; Mismatches 321; Indels 379; Gaps 32;

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable ubiquitin carboxyl-terminal hydrolase 1 (EC 3.1.2.15)
DE (Ubiquitin thiolesterase 1) (Ubiquitin-specific processing protease 1)
DE (Deubiquitinating enzyme 1).
CN UBPI OR SPCC16A11.12C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Guillermo R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gellens S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tilvey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Gadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Usery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announcement/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL109957; CAB53084.1; --
DR PIR; T41085; T41085.
DR GenDB_Spombe; SPCC16A11.12c; --
DR InterPro; IPR006615; DUSP.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00443; UCH; 1.
DR SMART; SM00695; DUSP; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS00235; UCH_2_3; 1.
KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT ACT_SITE 288 BY SIMILARITY.
FT ACT_SITE 798 BY SIMILARITY.
FT ACT_SITE 806 BY SIMILARITY.
SQ SEQUENCE 849 AA; 98655 MW; C8118042229BA9CB CRC64;

Query Match 8.1%; Score 693.5; DB 1; Length 849;
Best Local Similarity 21.3%; Pred. No. 5.6e-36;
Matches 233; Conservative 160; Mismatches 321; Indels 379; Gaps 32;

QY 534 PVKATSLTLEGRLKRTQQLNGRDYEMVPPVPRALYHYWG-ANLALPRPV--IKNSKT 590
DB 70 PITQWRLLDEKNEKHS--LEESIDYSVASLWMLVWVFGLEGIAERKVLVGLAAE 127
QY 591 DIPELELP-RVLLFLRQOPATRTQOSNIWNMGNV-SP-----NAPLKRVLAYTGCESR 644
DB 128 QKPFVDIYPINFTHLVFDP-----INGENTSPLYQIDEPYHSDEYAFSPSR 177
QY 645 MOTIKIEIHEYLSQRRIKE-EDMRLWLYNSEN-----YLTLLDDEHKL-----EY----- 689
DB 178 SDTLRSLYKQVMEAFQISDGTSLRWLYLNKSNLSRFSVLSSEFNDQPAIALLSAYAVCM 237
QY 690 ---LKIQDEQHLVIEVRKNDMSWPEMSFIANSKSIDRHKVPTKGTGATGLNLTGTCFVN 746
DB 238 IFEIDIADGS-LLEFQHPNGEWLSD-----SITKEQNLITNKEIGLCGLYNLNGSCYMN 291
QY 747 SSIQCVNSTQPLTOYFISGRHLYELNRNPIGMKGMHAKCYGDLVQEL--WSGTQKNVAPL 805
DB 292 SALQCMHITHETKYLKSDSYEKDINYNPLGMGKVALSYASLLKMIHTADHMSVSPS 351
QY 806 KLRWTIAYAPFNGFQODSOELLAFLLDGLHDLNRVHEKPYVELKDSGRPD----- 860
DB 352 SPKFIIGFNTYFSGYRQDSQEFIAFFLDGLHDLNRVHEKPYVELKDSGRPD-----RPLDFDEH 405
QY 861 ----NEVAAEAMDNLRNRSIVVDLFGQLRSOVKCTCGHISVRDPENFLSLPLMD 916
DB 406 PHVORVANQCWIITKNDISIIVLFGQMYKSTFECISYQKSTAFDFPFVTLPLPTS 465
QY 917 SYMHLEITVIKLDGT-TPVRYGLRLNMDKTYGLKKQLSD---LGLNSQQLLAENVHGS 972
DB 466 AKWRHKVYVVPFGTQSPVELYLLLMESTVQMKFQATEKLQKMGLEGELTACDIYRG 525
QY 973 NIKNPQDNQKRLSVSGFLCAFEIPVPVSPISASSTPTQDFSSSPSINFTLTNGDL 1032
DB 526 KYVKLNKNDKISKKI-----
QY 1033 PRPIFIPNGMNTVVPCTGKNTFTNGVNG-HMPSLPDPSPTGYIIAHRKMMRTLYFL 1091
DB 542 -----HKWDHVLVYGSTANGLTPIVHGCKRPAMPG----- 572
QY 1092 SQKQKPSLFGMPLIVCTVHTKKD-----LYDAVWIOVSRLASPLPQ 1136
DB 573 SYQSN--DVFGFPL---QLNVRNRVLTNDLVKEIVELYRYAGIDVAIGTLQGLKRM 626
QY 1137 EASNEAQDCCDSMGYQYPTLRVWQDGNSCAWCPYFCRGCKIDCGEDRAFIAGNAYIA 1196
DB 627 ESKAGNKEIKEIVK---RFEIVVEE-----EIV 653
QY 1197 VMDPTALHLRYQTQ-ERVDHEHSVQSRAQAPINLDSCLRAFTSEBELGENEMY 1255
DB 654 ID-DKTVIMCLMNDQYKELFYNCWEIPEKIQFMHESITLEDCLLEFSPKQLDQDSWY 712
QY 1256 CSKCKTHCLATKKLDLRLPPLIITHLAKRFQVNG---RWIKSQIKVFPRESPPSAPL 1312
DB 713 CPCKAFRPRATKLEIWRLPKILVILNRFPBGHGDLERRRKRRDLVVPYFDLNLKQFL 772
QY 1313 VPRDPALCQHKPLTPQGBELSEPRILAREVKVDAQSAGEEDVLLSKSPSLSANIIS 1372
DB 773 SP----- 774
QY 1373 PKGSPSSSRKSGTSCPSKSNKSNPSPRTLGRSKGRRLPQIGSKNKLSSKENLDASKE 1432
DB 775 ----- 774
QY 1433 NGAGQICELADALSRGHVLGSGQPELVTPQDHEVALANGFLYEACGNGSYNGQLNHS 1492
DB 775 -----FIKHEWLSQKS----- 787
QY 1493 EBDSTDQREDTRIKPIYINLYAISCHSILGGHVVYAKNPCK-WYCYNDSSCKELHP 1551

-- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
ubiquitin + a thiol.
-- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
Isoid=075604-1; Sequence=Displayed;
Name=2;
Isoid=075604-2; Sequence=VSP_005256, VSP_005257;
Note=No experimental confirmation available;
Note=Belongs to peptidase family C19.
-- SIMILARITY: Belongs to peptidase family C19.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF079564; AAC28392.1; .
EMBL; BC002854; AAH02854.1; .
EMBL; BC002955; AAH02955.1; .
MEROPS; C19.013; .
Genew; HGNC:12618; USP2.
MIM; 604725; .
GO; GO:0004197; F:cysteine-type endopeptidase activity; TAS.
InterPro; IPR001394; UCH-2.
Pfam; PF00443; UCH; 1.
PROSITE; PS00972; UCH_2_1; 1.
PROSITE; PS00973; UCH_2_2; 1.
PROSITE; PS02335; UCH_2_3; 1.
KW Ubiquitin conjugation pathway; Hydrolase; Thiol protease; Multigene family;
KW Alternative splicing.
FT ACT_SITE 24 24 BY SIMILARITY.
FT ACT_SITE 297 297 BY SIMILARITY.
FT ACT_SITE 305 305 BY SIMILARITY.
FT VARSPPLIC 1 252 Missing (in isoform 2).
FT VARSPPLIC 253 258 /FTID=VSP_005256.
FT CONFLICT 594 594 /FTID=VSP_005257.
FT CONFLICT 602 605 L->H (IN REF. 1).
FT CONFLICT 605 605 PSM -> TSPI (IN REF. 1).
SQ SEQUENCE 605 AA; 68071 MW; APP4DA9344D21812 CRC64;

Query Match 4.9%; Score 420; DB 1; Length 605;
Best Local Similarity 16.8%; Pred.No. 6.7e-19;
Matches 211; Conservative 86; Mismatches 259; Indels 698; Gaps 27;

333 HLLEDDYQINSVKVNLAFNLLFOVCHIVGLRPAATPEEGQIRGWLERSRYGLOA 392
21 HYAKSGYATPSSYGANLAASLLEKE---KLGFKPVPTS-----SFLTRPTVGPSS 70
393 GHNWFIIISMQWQWQVYKYDANPVVIEPSSVINGKYSFGTAHPMEQVEDRIGSSLS 452
71 -----LLDYDRGRPLLRP-DITGGGK-----RAESQTRGTERPLSGSL 108
453 YVNTTEKPSDNISTASEASEATAGSGFLYSATGADVCFARQNTSDNNQCILLGANGNI 512
109 -----GSGGFPYGVY-----NNCL-----S 123
513 LLHLNPKQGAIDNQPLVTPPV-----KATSLTLEGGRKRTPLIHGRD----- 558
124 YLPINAYDQGVTLTKLDSQSLARDFSSLTSDSYRIDPRNLGRSPMLARTEKELCTIQ 183
559 ---YEMVPEPWRLY-HWYGANLAPREVINKSTDIPELELPFVYLLFLRQOPATRTQQ 615
184 GLYTASCEPYLDVLYENYGR-----KGSASQVPS-----QAPPSR--- 219
616 SNIWVWNGVPSNAPLKRVLAVTGCFSRMQTIKEIHEVLSORLRIKEEDMRLLWYNSEN 675
220 -----VPEIISP-----TRPIGRY-----TLW----- 237
676 YLTLLDDHDKLEYLKIQDEOHLVIEVRNKMGMPEMSFIANSKIDRHKVPTKGTATG 735

238 -----ETGKGQAFPSRSS-----SPGRDGWNSKSAQGLAG 268
736 LSNLGNCTCFMNSSIQCVSNTQPLQYFISGRHLYELARTNPIMGKGMHA--KCYGDVLQVE 793
269 LRLNGNCTCFMNSILQCLSNTRRLRDYCLQRLYMRDLHH---GSNAHTALVEEPFKLIQT 324
794 LWSGTQKN-VAPLKLRTWIAKYAPRENGFOQDSQELLAFLLDGLGHEDLNARVHEKPYVEL 852
325 IWTSPNDVSPSEFKTQIQRYAPRFVGYNQDQAEFLRFLDGLGHNEVRKVTURPKSNP 384
853 KDSGRPDWEVAAEAWNDHNRNRSIVVDLPHGQLRSQVRCCTGCHISVRPDPFNFLSLP 912
385 ENLDHLDPDEKGRQWRYLERSDRIGDLFVQGLKSLCTDTCGCGYSTVPDPDWLISLP 444
913 LPMDSYMHLEITVILKIDGTPVRYGLNMDKVTGLKKQLSDLCGLNSQLLAELVHGS 972
445 IAKRGY----- 450
973 NIKNPPQDNQKRLSVSGFLCAFEIPVPSFISASPTQTDGSSSPSTNEMFTLTNGDL 1032
451 ----- 450
1033 PRPIFNGMNTVVPCTEKNFTNGMVGMPSLDPSPTGYIIAHRKMRTELIFLS 1092
451 ----- 450
1093 SQKNRPSLFGMPLIVPCTVHTRKXLDYAVMIQVSRLASPLPPOEASNHAQDCDDSMGYQ 1152
451 -----P3----- 452
1153 YPFLRVVQKDGNSCAWCPWYRPCRCKIDCGEDRAFIGNAYIAVDWDPTALHLRYTSQ 1212
453 ----- 452
1213 ERVDEHESVQSRAQAEPLNLDSCLAFTSEBELGENMYCCKTHCTLA TKLDDLW 1272
453 -----VTLMDCMRLFTKEDVLDGDEKPTCCRCGRKRCIKFESIQ 492
1273 RLPILIIHLKRFQVNGRWIKSOKIVKFPRESPPSAFLVPRDPALCOHKLPTQGDDEL 1332
493 RFPKILVLHLKRFSESRIKSTLTFVNFPLRLDL----- 527
1333 SEPILAREVKVDQAQSSAGEEDVLLSKSPSSLSANTISSPKGSPSSRSKSGTSCPSKN 1392
528 ----- 527
1393 SSPNSPRTLGRSKRLRLPOIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLG 1452
528 -----LREFASENT----- 536
1453 GSQPELVTPQDHEVALANGFLYEHAACNGYNGQLGNHSEEDSTDQREDTRIKPIYNL 1512
537 -----NHA-----VYNL 543
1513 YAIISCHSGLGGHVTYAKNPNC-KWCVNDSSCKELHPDEIDTDSAYILFYE 1565
544 YAVSNHSQTTWGGHTAYCRSPGTGEWHFTNDSSVTPMSSSQVTSAYILFYE 597

RESULT 15
USP2_MOUSE
ID USP2_MOUSE STANDARD; PRT; 353 AA.
AC O88623;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
thiolesterase 2) (Ubiquitin-specific processing protease 2)
DE (Deubiquitinating enzyme 2) (41 kDa ubiquitin-specific protease).
GN USP2 OR USP41.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Gong L., Yeh E.T.H.;
 RT "Cloning and expression of the human and mouse UBP41";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL; AF079565; AAC28393.1; -;
 DR MEROPS; C19.013; -;
 DR MGD; MGI:1858178; Usp2.
 DR InterPro; IPR001394; UCH-2.
 DR Pfam; PF00443; UCH; 1.
 DR PROSITE; PS00972; UCH_2.1; 1.
 DR PROSITE; PS00973; UCH_2.2; 1.
 DR PROSITE; PS02335; UCH_2.3; 1.
 KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT_SITE 24 24 BY SIMILARITY.
 FT ACT_SITE 297 297 BY SIMILARITY.
 FT ACT_SITE 305 305 BY SIMILARITY.
 SQ SEQUENCE 353 AA; 40581 MW; 4FFB39A225FE8F11 CRC64;
 CC -----
 Query Match 4.8%; Score 412; DB 1; Length 353;
 Best Local Similarity 16.9%; Pred. No. 9.3e-19;
 Matches 144; Conservative 59; Mismatches 134; Indels 516; Gaps 10;
 QY 717 ANSSKIDRHKVPTEKGATGLNLTGNTCFMNSSIQCVSNTPQLTQYFISGRHLYELNRTNP 776
 DB 5 AKNSK-----SAQGLAGLRNLGNTCFMNSILOCLSLNTRLDYCLQRLMYRDLGHTS- 56
 QY 777 IGKMGHMA--KCYGLVQELMSGTQKN-VAPLKLRTWTIAKYPFRNGFQOQDSOELLAF 833
 DB 57 ---SAHTALMEEFAKLIGTWTSSNDVVSSEFKTQIQRYAPFRMGYNQQAQAEFLRFL 113
 QY 834 LDGLHEDLNVRHEKPYVELKDSGRPDWEVAEAAWDNHLRNRNRSIVVDLFHGQLRSQVKC 893
 DB 114 LDGLHNEVNRVAAREKASPETLDHLDEKGRQWRKYLERSRIGDLFVGQLKSLTC 173
 QY 894 KTCGHISVRFPDPFNFLPLPMDSTMHLEITVTKLDGTPVRYGLRLNMDEKYTGLKKQL 953
 DB 174 TDCGYCSTVDFDFWDLSPFAKGY-----198
 QY 954 SDLCGLNSEQILLAEVHGSNIKPNFPQDNQKVRLSVSGFLCAFEIPFVSPISASSPTQTD 1013
 DB 199 -----198
 QY 1014 FSSSPSTNEMTLITNGDLPRPIPIPGMNTVVPCTGTEKNTGNGVNGHMPSLDPSPT 1073
 DB 199 -----198
 QY 1074 GYIIAVHKKMNRTELYFLSSQKNRPSLFGMPLIVPCTVHTRKLDYDAVMIQVSRLASPL 1133
 DB 199 -----198
 QY 1134 PPQASNEAQDCDDSMGYQYPTFLRVVQKGNCSAWCPWYRCGCKIDCGEDRAFTGNA 1193
 DB 199 -PE-----200
 QY 1194 YIADVWDPTALHLRYQTQSERVVDHESVEQSRRAQAEPIINLDSCLRAFTSESELGENEM 1253
 DB 201 -----VTLMDCMRLFTKEDILDGDEK 221

QY 1254 YYCSKCKTHCLATKLDLWELPILIIHLKRQFVNGRWIKSQKIVKFPRESFDPSPALV 1313
 DB 222 PTCRCRAKRCIKKFSVQRFKILVHLKRFSESRIRTSKLTTFVNFPLRDLD-----275
 QY 1314 PRDPALCOHKPLTPQGDSEPRILAREVKVDAOSSAGEEDVLLSKSPSSLSANIISP 1373
 DB 276 -----275
 QY 1374 KGSFSSSRKSGTSCFSSKNSPNSPRTLGRSKGRLLRPLQIGSKNKLSSKENLDASKEN 1433
 DB 276 -----LEEFASENT-----284
 QY 1434 GAQICELADALSRGHVLSGSPQLVTPQDHEVALANGFLYEHEACNGYNSGQLGNHSE 1493
 DB 285 -----NHA- 287
 QY 1494 EDSTDQREDTRIKPIYNIYALISCHSGLIGGGHYTYAKNP-NCKWYCVNDSCKELHPD 1552
 DB 288 -----VYNLYAVSNHSGTTMGGHYAYCRSPVTGWHHTFNDSSVTPMSSS 332
 QY 1553 EIDTDSAYILFYE 1565
 DB 333 QVRTSDAYILFYE 345

Search completed: February 11, 2004, 15:39:17
 Job time : 32 secs